HAVE THE JAPANESE BECOME HAPPIER OVER THE LAST 50 YEARS?
Healing broken time-series data

Tineke DeJonge[1] Akiko Kamesaka[2] & Ruut Veenhoven[3]

EHERO working paper 2019/3

ABSTRACT

Periodical survey studies show that average happiness has risen in most developed nations since the 1970s, similar survey data in Japan do not show such a clear trend. Average happiness in Japan was first assessed in 1958, and has since been assessed in at least six survey programs. The questions used in these surveys are not identical and for this reason it is difficult to identify a time trend in the data. In this paper, we apply four methods for homogenizing these data, each of which allows computation of an average on a numerical scale ranging from 0 to 10: 1) simple ordinal ordering of response options by their ranks, starting with 1 for the least happy response, followed by linear stretch to scale 0-10, 2) assigning fixed weights to particular response options, such as 8 for 'happy', 3) assessing weights in the context of the response scale, using native speakers as judges, and 4) deriving weights of verbal response options using a reference distribution obtained in the same year from answers to a similar survey question rated on a numerical response scale. We find that the Japanese have not become any happier over the last 50 years.

Keywords: happiness, life-satisfaction, trend analysis, response scale homogenization

1 INTRODUCTION

There is a rising interest in happiness and in particular in average happiness in nations. One question in this context is whether citizens are getting any happier. This appears to be the case for most modern countries since the 1970s (Veenhoven 2014), but it is as yet unclear whether happiness has also increased in Japan. Though happiness has been assessed on a regular basis in Japan since 1958, the use of different survey questions clouds the view on the time trend. In this paper we present techniques for dealing with heterogeneity in survey questions on the same topic, drawing on DeJonge et al (2017), and apply these techniques to the

[1] J.J. DeJonge, Erasmus University Rotterdam, The Netherlands, Erasmus Happiness Economics Research Organization, e-mail: dejonge@ese.eur.nl.
[2] A. Kamesaka, Aoyama Gakuin University, Japan, School of Business Administration and Former Visiting Research Fellow, Economic and Social Research Institute (ESRI), Cabinet Office, Government of Japan, Japan.
[3] R. Veenhoven, Erasmus University Rotterdam, The Netherlands, Erasmus Happiness Research Organization and North-West University South-Africa, Opentia Research Program.

heterogeneous Japanese survey questions on happiness[4].

1.1 Interest in Happiness[5]

The high ranking of happiness in value hierarchies of students all over the world reflects the human desire for a person to have a satisfying life for themselves and their children (Diener and Oishi 2004). Individually, people seek ways to a more satisfying life and in Western societies this quest manifests in the soaring sales of 'how-to-be-happy books', such as 'The art of happiness' by the Dalai Lama (1998). It also reflects in the development of life-coaching businesses.

Citizens in western societies call on governments to improve their social conditions which they see as a requirement for their happiness and 85% of the British agree with the statement that 'a governments prime aim should be achieving the greatest happiness of the people, not the greatest wealth' (BBC 2006, question 14). Consequently, interest in happiness is rising among policy makers. Happiness is a new topic on the political agenda, next to sustainability. The international conference on Happiness and Wellbeing held at the UN headquarters in New York in April 2012 is a recent manifestation of this trend. This conference was followed in June 2014 by a decision of the UN General Assembly to celebrate an 'International Day of Happiness' on March 20th every year (Thinley 2012),

1.2 Research on Happiness in Nations

This call for greater happiness has instigated a lot of research on the level of happiness in nations. Happiness is commonly measured using single direct questions, such as: Taking all together, how happy would you say you are these days? Would you say you are 'very happy', 'pretty happy' or 'not too happy' these days?

The level of happiness in a nation is typically characterized by the average response to such questions in a general population survey. All research findings on average happiness in nations are gathered in the World Database of Happiness (WDH), Collection of Distributional Findings in Nations (Veenhoven 2018b). In October 2019 this collection contained 8,846 such observations on average happiness in a particular nation and year. Happiness has been assessed in almost all nations of the present-day world. A notable exception is North Korea.

The happiness data reveal wide differences in average happiness across nations. The lowest mean on a 0-10 scale observed was 1.6 in the Dominican Republic in 1962 and the highest, 8.5 in Costa Rica in 2007. Average happiness can be compared across time for some 20 nations. Trend data on nations for which at least 10 data-points over at least 20 years are available, are gathered in the 'Trend report on average happiness in nations 1946-2015' (Veenhoven 2018c). These trend data are used in the ongoing discussion about the quality of modern society. One topic in this discussion is whether societal modernization fits with human nature. If

[4] With the term 'happiness' we refer to the subjective enjoyment of one's life as a whole, a definition in which 'happiness' is synonymous with 'life satisfaction' (Veenhoven 1984).

[5] This section draws on Veenhoven (2018a, p.1).

not, one can expect that life is getting less satisfying in modernizing societies (Veenhoven & Berg 2013). A related issue is the effect of economic growth on average happiness in nations. In this context Easterlin (1974) argues that economic growth has typically not been accompanied by a rise in average happiness in developed nations over the past decades. This counter-intuitive finding has attracted much attention and manifests, for example, in the 7.260 hits[6] on "Easterlin Paradox"[7] in Google Scholar.

1.3 The case of Japan

Japan is an interesting case in the above-mentioned discussions on the effects of societal modernization and economic growth on average happiness in nations. Japan exemplifies a case of late but rapid modernization in the late19th century and in the second half of the 20th century, the country witnessed unprecedented economic growth. This encouraged several researchers to analyze the time trend in happiness in Japan.

Data on average happiness in Japan

Questions on happiness have figured in survey studies among the general population in Japan since 1958. The WDH contained 160 observations on average happiness in Japan in October 2019 covering the period from 1958 to 2016. These observations were measured using among others the following periodical survey programs:

- World Values Survey
- Lifestyle Preference Survey
- Life in Nation Survey
- Future Life Survey[8]
- East Asia Values Survey
- Japanese National Character Survey

Questions used for assessing happiness in Japan

Different questions on happiness have been used in these survey studies and some have slightly changed the wording of their questions or the labels of the response options over the years. The findings on average happiness in Japan have been obtained using over 30 different questions. These questions can be sorted into 5 groups of 'equivalent' questions, that is, questions that address the same kind of happiness, e.g. questions in which 'life-satisfaction' is the keyword. Comparison across time makes sense only for responses to such equivalent questions (Veenhoven 1994, section 7/2.4).

[6] Assessed on 18-10-2019.
[7] The Easterlin paradox states that at a point in time happiness varies directly with income both among and within nations, but over time happiness does not trend upward as income continues to grow, https://en.wikipedia.org/wiki/Easterlin_paradox, assessed 18-10-2019.
[8] The official name of this Japanese Government's survey is "Public Opinion Survey on the Life of the People", https://www.gov-online.go.jp/eng/pdf/summaryk16.pdf.

Equivalent questions of happiness in Japan are available for the following measures of happiness:

- 25 single questions about 'Happiness', 1975-2012
- 90 single questions about 'Life-satisfaction', 1958-2012
- 17 scores on an affect balance scale[9], 2006-2016
- 15 single questions about present best/worst possible life (Cantril ladder), 1962-2016
- 13 single questions about 'Feeling happy', 1972-2010

All of these questions have a discrete number of response options, varying from 2 to 11, but in the analyses the mean life satisfaction measured is treated as a continuous variable. Scholars use different approaches in studies on the trend in mean life satisfaction in Japan to deal with this discrepancy. These different approaches lead to different conclusions, which confuses the discussion on how happiness develops in Japan.

The longest time series on average happiness in Japan results from the Life in Nation (LIN) survey, with 58 measurements over the period 1958-2013. The questions used in the LIN survey and the changes in the wording of the labels of the response options are central to the divergence in views on the trends in mean life satisfaction in Japan.

1.4 Aim and plan of this paper

In this paper we will consider several approaches and methods to obtain the trend in Japanese life satisfaction with the aim to remove the clouds that obscure the direction in which this trend evolves, using the data taken from the LIN survey for the period 1958-2007 resulting from different approaches.

We will first look in more detail at the question on life satisfaction used in the LIN survey and the trend resulting from the measurements using the straightforward Rank Method in Section 2. Next, in Section 3, we will discuss the results taken from three papers paying attention to time trends in happiness in Japan in relation to economic development making use of the time series from the LIN survey. Following on this discussion, we will consider four different methods for scale homogenization[10] that can be used to transform the mean life satisfaction values of the different LIN surveys into values on a numerical scale from 0 to 10 in Section 4. Each following method is more advanced than the previous one. The aim of applying these different scale homogenization methods is to obtain a realistic view on the direction in which the trend in Japanese life satisfaction evolves. In Section 5 we will present the resulting trend in happiness in Japan when applying each of these methods. A discussion on the results we find, is given in Section 6. We end with conclusions in Section 7.

[9] An affect balance scale consists of a number of positive and negative mood states and respondents must report if they have been in each of these states in the last period of time, e.g. week.

[10] The descriptions of the scale homogenization methods are based on those written by us in previous papers and books, among others (DeJonge et al 2014; DeJonge et al 2017).

2 LIFE SATISFACTION OVER TIME MEASURED IN JAPAN USING THE LIN SURVEY

2.1 Life satisfaction in the LIN Survey

The measurements on life satisfaction in the LIN survey started in 1958. Since then, average life satisfaction has been measured almost every year in Japan and in some years even twice. The measurement is done using a leading question with a 4-point verbal response scale.

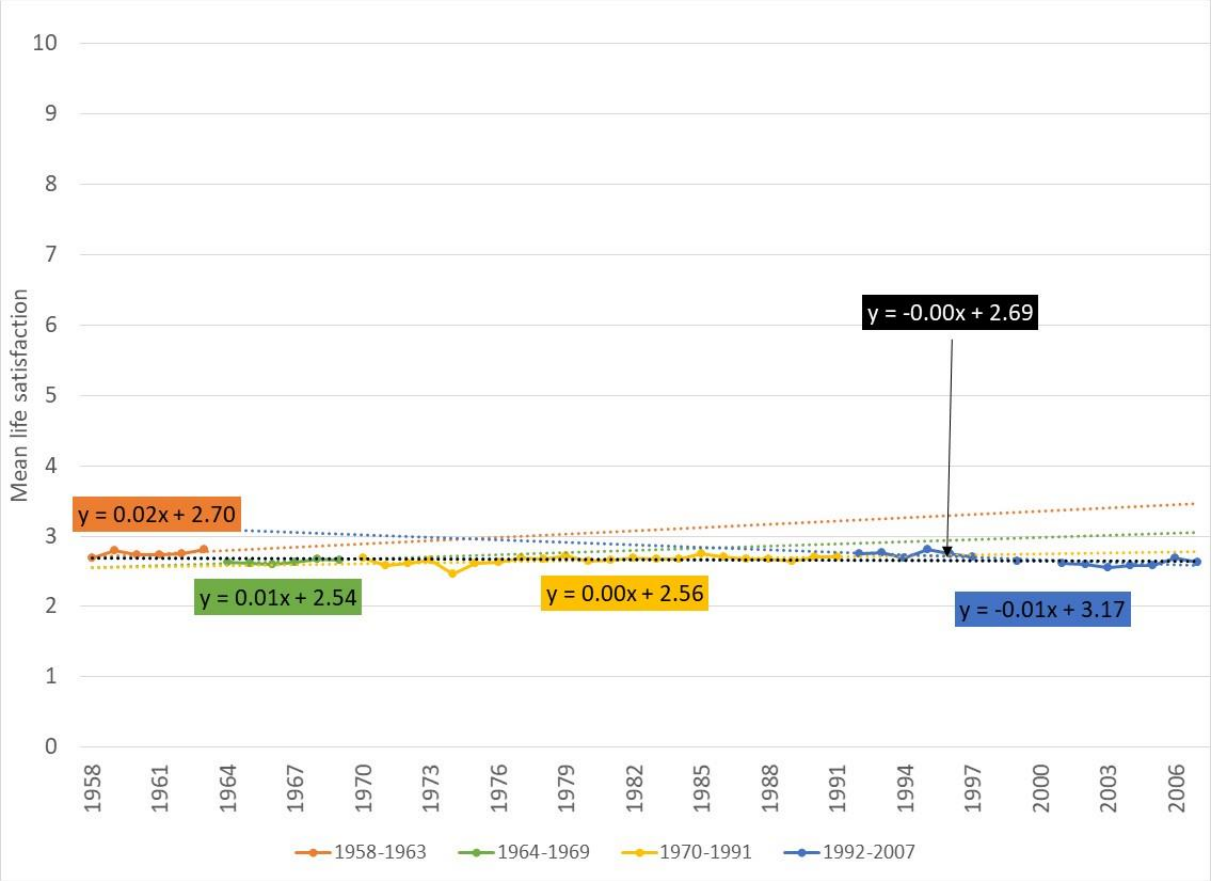
How respondents interpret survey questions and the corresponding response options depends on different factors of which the wording used to label the options is an important one. Suzuki (2009, pp. 84-85) remarks that the labels of the response options of the question on life satisfaction in the LIN survey have significantly changed between 1958 and 2007. In 1964 the labels changed from long descriptions to more tersely formulated statements which translated into English read 'Extremely dissatisfied', 'Fairly dissatisfied', 'Rather satisfied, but not sufficiently' and 'Sufficiently satisfied'. A second change happened in 1992 when the scale was made more symmetric with the introduction of the labels 'Dissatisfied', 'A little dissatisfied', 'Rather satisfied' and 'Satisfied', and a don't-know choice 'I cannot say either satisfied or dissatisfied' was added. Just like Suzuki, Stevenson and Wolfers (2008) point to the changes in the survey question for measuring life satisfaction between 1958 and 2007 in terms of the labels of the response options. Stevenson and Wolfers (p. 21) make one additional split to Suzuki's in the time series on life satisfaction to account for a change in the focus of the leading question on "life at home" until 1969 to general life satisfaction in 1970. Suzuki also mentions this change (p. 8, footnote 2), but he, as a native speaker, is of the opinion that the question employed until 1969 should be interpreted as focusing on your life which makes the effect of a change less relevant.

2.2 Trend in life satisfaction in Japan based on the Rank Method

To be able to calculate statistics, such as mean and correlations, it is necessary to express the verbal labels of the response options as numerical values. The common practice is to assign ranks to the verbal response options of a discrete scale. In the rank method the sample mean is calculated as the weighted average of the ranks of the response options using the relative frequencies as weights. The resulting mean is a value within the range of the ranks.

In the case of the LIN question a value between 1 and 4. The trend in mean life satisfaction when applying the rank method to the time series from the LIN survey, is given in Figure 1. The time series is split into four time periods in Figure 1, based on the years in which the survey question was changed. The linear trend is given for each of these time periods and for the entire period 1958-2007. It can be seen from Figure 1 that mean life satisfaction in Japan fluctuates between 2.47 and 2.82 when using the ranks of the response options for calculating the mean.

Figure 1
Mean life satisfaction in Japan, 1958-2007, Rank Method



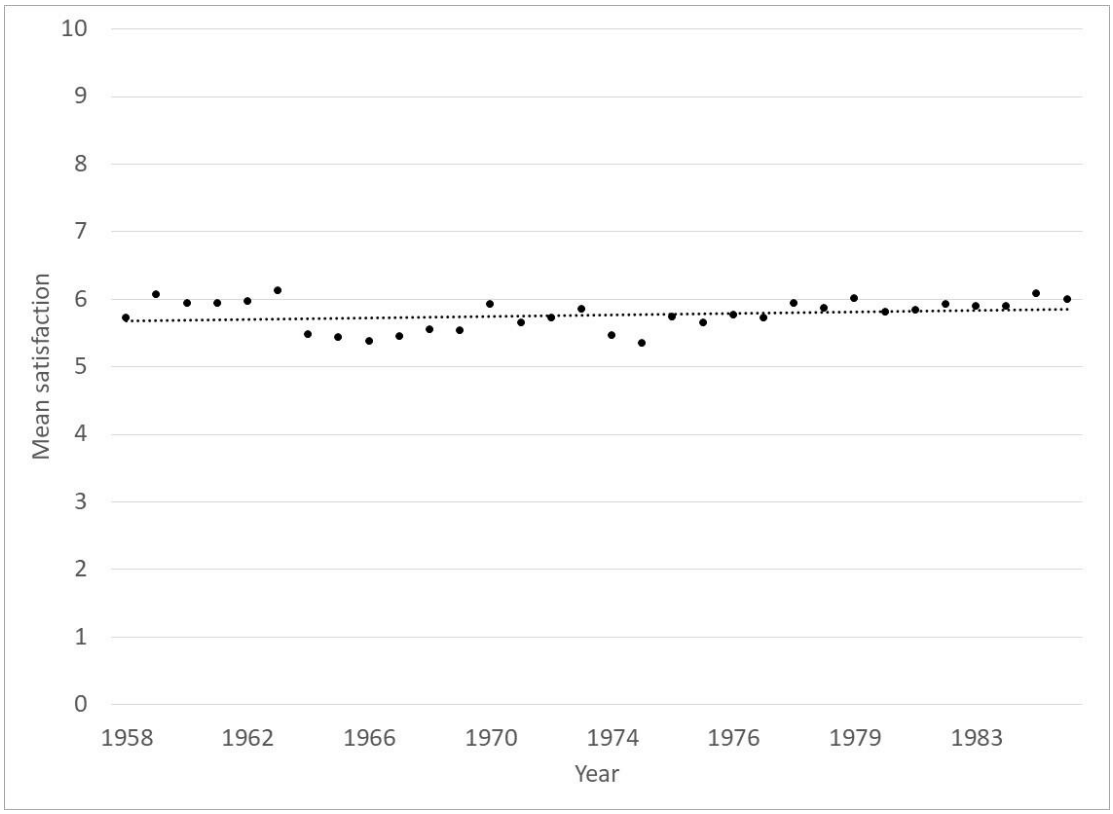
3 DIVERGENT VIEWS IN THE LITERATURE ON THE TREND IN HAPPINESS IN JAPAN

We will discuss the approaches chosen to analyze time trends in happiness in Japan in relation to economic development described in papers by Easterlin (1995), Stevenson and Wolfers (2008) and Suzuki (2009). All these authors have used the time series on life satisfaction from the LIN survey. The analysis by Easterlin is restricted to the period 1958-1987. The other authors have also look to the period the follows until 2007.

3.1 Analysis by Easterlin

Easterlin (1995) shows a five-fold multiplication of real per capita income in Japan in the period from 1958 to 1987. The level of living in Japan was propelled from that of about one-eighth the United States at the beginning of that period to about two thirds some thirty years later. Yet, against expectations, the average level of life satisfaction based on measurements using the question on life satisfaction from the LIN survey was unchanged as can be seen from Figure 2.

Figure 2
Mean life satisfaction in Japan, 1958-1987



Note: mean values in Figure 2 taken from the WDH [11] and used in Figure 3, Easterlin (1995), p. 40

[11] https://worlddatabaseofhappiness.eur.nl/hap_nat/nat_fp.php?cntry=6&name=Japan&mode=3&subjects=185&publics=19, Measure type: 121 C 4-step verbal Life Satisfaction.

The mean values in Figure 2 are the result of a weighted average approach. In this approach mean values are calculated as the weighted average of fixed values on scale 0-10 assigned to the response options of a verbal scale question using the relative frequencies measured in a wave of surveying as weights. We will refer to this method as the Thurstone Conversion to which we will come back in Section 4.2. The regression line fitted to the data in Figure 2 applying an ordinary least squares regression has a coefficient that is not statistically significant. Easterlin concludes that the results from the LIN survey signal that there is no significant correlation between national income and subjective well-being in Japan.

3.2 Analysis by Suzuki

Easterlin's work inspired Suzuki (2009) to delve deeper into the labelling of the response options of the question on life satisfaction from the LIN survey. Suzuki argues that the changes in the wording used to label the response options (see Section2.1) have an effect on the response option a respondent may choose in terms of the ranks of the response options. He states, for example, that it is reasonable to expect that someone who would choose the option with rank 2 when asked before 1963, would choose the option with rank 1 after the change of 1964, and that this would lead to a significant decrease in the measured mean value. Suzuki underpins this argument by presenting the figure shown in Figure 3 with a 3-point moving average in mean life satisfaction resulting from the Rank Method which differs from the weighted average approach as the ranks of the response options are used as weights instead of fixed values being assigned to these options to calculate the mean value[12].

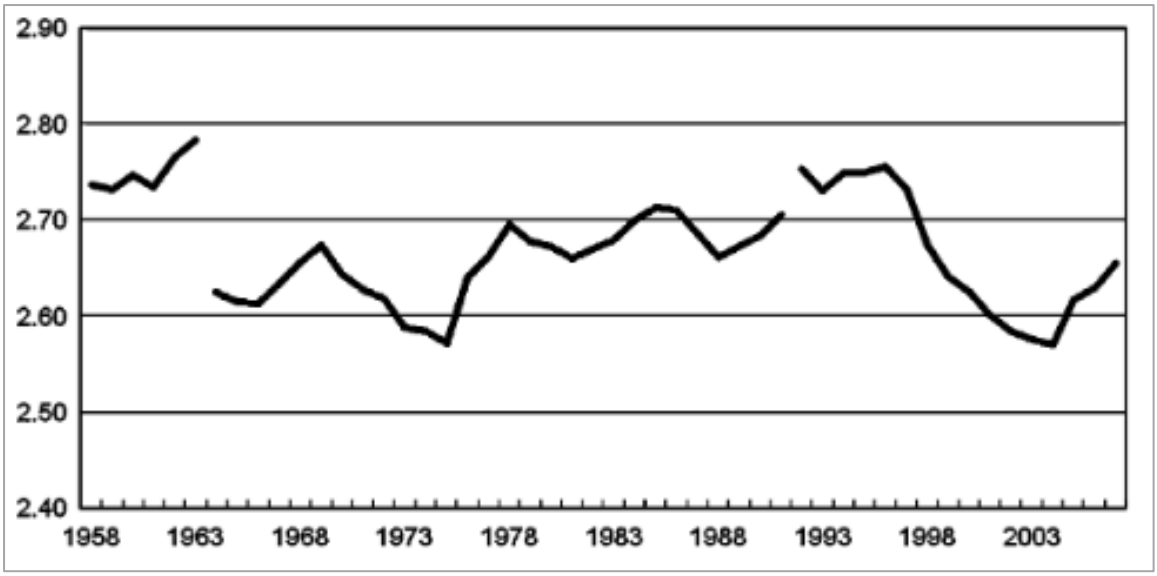
In response to the change in the labelling of the response options Suzuki initially set up a regression model for examining the correlation between national income per capita and life satisfaction, including two dummy variables for the periods 1958-1963 and 1964-1991 to indicate the different periods of measuring life satisfaction. This model resulted in a very weak, though significant correlation (R^2 =0.272) between changes in national income and life satisfaction over time.

Given the trend in mean life satisfaction, Suzuki reasons that life satisfaction may not only be affected by national income per capita but may also be affected by national income's per capita growth rate. In addition, Suzuki describes how many Japanese experienced a serious loss of confidence in their traditional economic system in the late 1990's, which he took into account in the regression model as a dummy variable for structural depression. The final model with national income per capita, the growth rate and the three dummy variables as independent variables and life satisfaction as dependent variable shows a strong association for the period 1958-2007 (R^2 =0.670). Suzuki's final conclusion can be summarized as that it is too short-sighted to state that there is no significant correlation between national income and subjective well-being on the basis of the LIN survey, since additionally taking into account the changes in the labelling of the response options, the national income's per capita growth rate and the loss of confidence in the traditional

[12] The weighted average approach is a generalization of the rank method.

economic system reveals that the level of life satisfaction of Japanese people is affected by the economic development of their country.

Figure 3
Three-point moving average in life satisfaction in Japan, 1958–2007



Note: original Figure 2, Suzuki (2009), p. 85[13]

3.3 Analysis by Stevenson and Wolfers

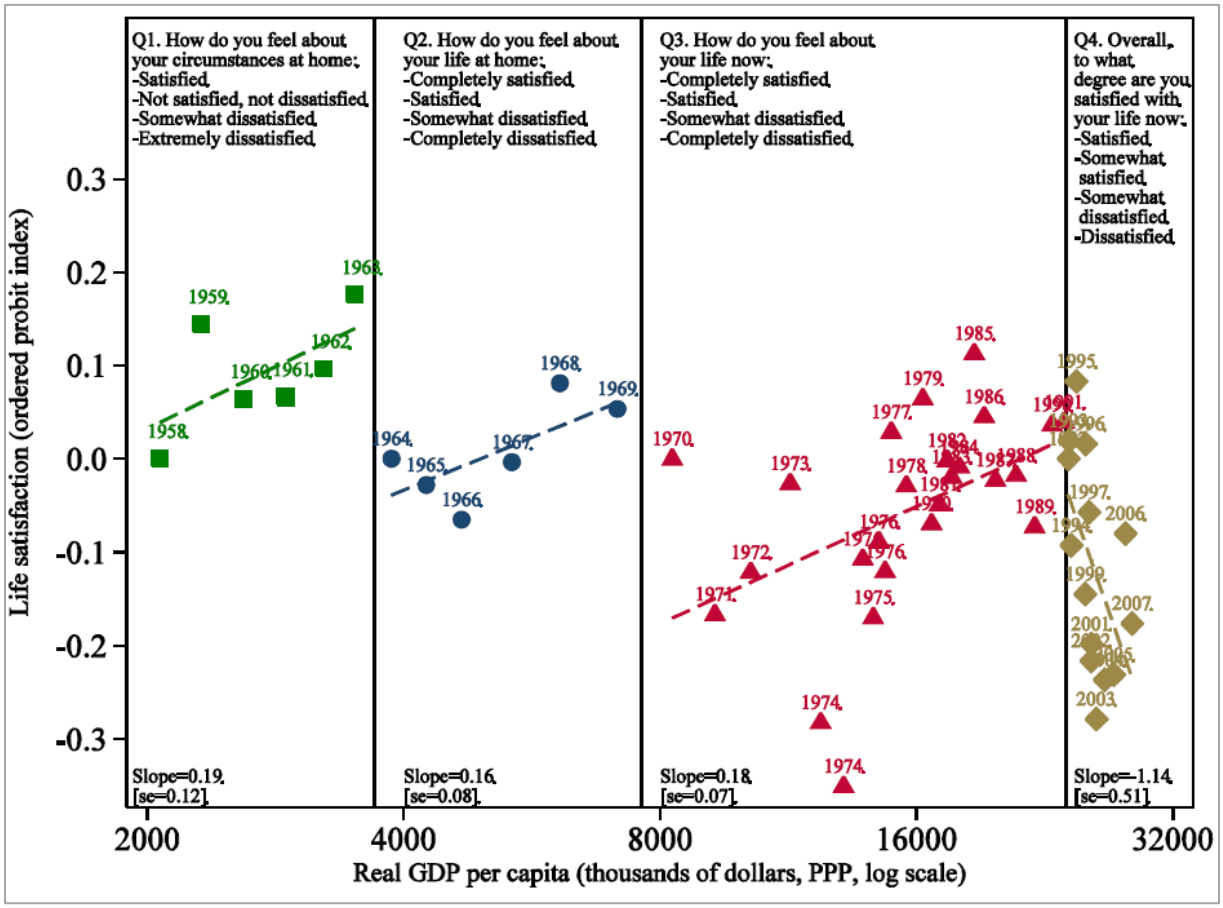
Stevenson and Wolfers (2008) have also questioned the conclusion of Easterlin that there is no significant correlation between national income and subjective well-being in Japan given the results of the LIN survey. Stevenson and Wolfers state that due to the changes in the life satisfaction question, the time series based on each question should be presented separately. They then performed an ordered probit analysis of subjective well-being on time fixed effects and GDP per capita for each of these subseries, getting the results shown in Figure 4.

Stevenson and Wolfers conclude from Figure 4 that throughout the period in which Japan moved from poor to affluent, the first three panels, subjective well-being rose with GDP per capita and that, since 1992, the Japanese economy has shown very little growth, which has come with a sharp fall in average life satisfaction. They also noted that three episodes in the economic progress could be distinguished in Japan during the period of interest: strong growth and decreasing unemployment rates from 1958-1969, slower growth and increasing unemployment rates from 1970-1991 and even slower growth from 1992 onward coinciding with the emergence of large scale unemployment. These episodes correspond well to the changes in the survey question for measuring life satisfaction. In a second step Stevenson and Wolfers combined all the data on life satisfaction and GDP in a regression model with mean life satisfaction per wave as the dependent variable and three dummy variables

[13] Suzuki restricts the range of the vertical axis from 2.40 to 2.90, instead of a range from 1 to 4. This makes the peaks and dips looks steeper than they actually are.

defining the periods bounded by the changes in surveying, the unemployment rate and the log of GDP per capita. The ordered probit estimation of the model resulted in negative coefficients for the dummies and the unemployment rate and a positive coefficient for GDP per capita. This led to their conclusion, which contrasts that of Easterlin: there has been a strong relationship between life satisfaction and GDP per capita growth in Japan since 1958.

Figure 4
Life satisfaction and GDP per Capita over time in Japan[14]



Note: original Figure 18, Stevenson and Wolfson (2008), p. 69
Source: Life in Nation surveys, 1958-2007

3.4 Reflection on the three papers

The above discussion makes clear that the views on the trends in life satisfaction in Japan based on data taken from the LIN survey and how these trends relate to trends in economic indicators diverge, depend on how the Japanese data is treated and interpreted.

Easterlin (1995), Suzuki (2009), and Stevenson and Wolfers (2008) all studied the trend in mean happiness in Japan using Japanese LIN data, in relation to trends in economic indicators. The flat regression line for mean happiness per year on a

[14] Stevenson and Wolfers used short labels for the response options in Figure 3, instead of the longer formulation which they describe at pp. 21-22 of their paper.

scale from 0 to 10 when applying a Thurstone Conversion (see Section 4.2) and the rapid growth of the national income, brings Easterlin to the conclusion that economic growth has not been paralleled by a raise in happiness in Japan. Suzuki follows a different approach. He regresses mean happiness on a scale from 1 to 4 obtained by applying the Rank Method on the national income, the growth rate, the loss of confidence in the traditional economic system and two dummies to indicate a change in the question used for measuring happiness in the Japanese LIN survey. This makes him conclude that life satisfaction in Japan is affected by economic development. Stevenson and Wolfers go one step further: applying an ordered probit model with happiness as dependent variable and log GDP per capita, unemployment and, like Suzuki, dummies for the different questions used in the LIN as independent variables. They conclude that there is a strong relationship between life satisfaction and GDP per capita.

The method used to derive mean happiness from the LIN data is different in each of the three papers. We will now give some comments on the methods used by Suzuki and by Stevenson and Wolfers. Comments on the Thurstone Conversion implicitly used by Easterlin are given in Section 4.2, where we dive deeper into this scale homogenization method.

Disadvantages of the Rank Method

Some serious disadvantages of the Rank Method (see Section 2.2.) are that the range of the scale depends on the number of response options, the response options are considered to be equidistant which may not be realistic and finally, the verbal labels of the response options are ignored. In addition it is assumed that the topic of interest, in this case life satisfaction in Japan, has a discrete distribution within a population.

Using dummies

A remark we want to make about the use of the dummies included in the models set up by Suzuki and Stevenson and Wolfers, is that these dummies are intended to account for the changes in the survey question, but in fact account for all the changes that took place in the period in which one of the questions was used. These dummies thus also represent the economic development in Japan in the period they cover and the developments taking place in society, such as the divorce rates, fertility rates and so on. All the trends in other things than the economic indicators in a certain period may affect mean happiness, both positively and negatively.

Distribution of the latent variable in ordered probit

In ordered probit it is assumed that the latent variable is normally distributed within a population (Stevenson and Wolfers 2008, p. appendix-1; Daykin and Moffat 2002, p. 161). We doubt whether this is a realistic assumption in general, as in the western world most people have a positive perception of their own well-being which makes the distribution of responses to questions on happiness skewed, with a long tail on

the left that represents ‘negative’ outcomes (Diener and Diener 1996; Cummins 2003; Frijters et al 2008).

Estimation of the ‘cut points’ in ordered probit models

The cut points estimated in ordered probit models are parameters used to model the relation between the observed, discrete variable y and the latent, continuous variable y*. In the case of the Japanese LIN question, three cut points k1, k2 and k3 are estimated assuming that:

y = 1 if -∞ < y* < k1
y = 2 if k1 < y* < k2
y = 3 if k2 < y* < k3
y = 4 if k3 < y* < ∞.

Given this relationship, the ordered probit model is based on the assumption that y_i^* depends linearly on x_i according to the following (Dayking and Moffat 2002, p. 160):

$$y_i^* = x_i' \beta + u_i, \text{ where } i = 1,\ldots,n \text{ and } u_i : N(0,1)$$

In the model of Stevenson and Wolfers y* is latent happiness and the scalar x’ß consists of log GDP per capita, the unemployment rate and the dummies for the different questions used in the Japanese LIN survey for measuring happiness. As we understand, parameters have to be estimated in ordered probit on both sides of the equation to make the model fit: the coefficients of the variables on the right side of the equation and the cut points on the left side of the equation. This gives the impression that the cut points depend on the values of the independent variables. The paper by Stevenson and Wolfers, however, does not submit substantiating evidence on this. This requires further investigation, which goes beyond the scope of this paper.

Interpretation of the values of the cut points

According to Daykin and Moffat (p. 162), the values of the cut-points estimated in ordered probit models are rarely given interpretation. They mention two suggestions for the interpretation of cut points. Firstly, they expect that the dispersion of the cut points depends on the extent to which most people strongly agree or strongly disagree with a statement. Secondly, they believe that the cut points adjust according to the wording of the statement: they expect that the more difficult the wording is to understand, the more the middle cut points might be apart from each other. In our opinion, however, the cut points should be interpreted as the transition points between two consecutive response options, and therefore it is doubtful that they can be dependent on the independent variables in the model. We wonder whether the strong growth in happiness reported by Stevenson and Wolfers is mainly to be attributed to this interdependency between the cut points and the independent variables in ordered probit, which forces happiness in the model to follow the

developments in log GDP per capita and the unemployment rate linearly as good as it gets. It would be a nice exercise to apply ordered probit to happiness and other indicators than log GDP per capita and the unemployment rate, to see if this would give the same cut points. If so, this would be an indication that the cut points can be considered as the transition points between two consecutive response options. If the exercise results in different cut points, this would be an indication that they depend on the independent variables used in the model.

Value range of the cut points

In addition to the above, the cut points estimated in ordered probit are hard to interpret, as they can take all values on range [-∞, ∞] . It would be easier to interpret these points if they are considered to be the transition points between response options on the continuum from 0 and 10.

In our opinion when happiness is measured using different survey questions, before drawing any conclusions the results should be homogenized before a trend can be identified. With this in mind, we will continue discussing four different scale homogenization methods that can be used to transform the mean life satisfaction values of the different LIN surveys into values on a numerical scale from 0 to 10. Each following method is more advanced than the previous one. The aim of applying these different scale homogenization methods is to obtain a realistic view on the direction in which the trend in Japanese life satisfaction evolves.

4 SCALE HOMOGENIZATION METHODS

We used four methods to homogenizing our Japanese data, each of which allows computation of an average on a numerical scale ranging from 0 to 10: 1) simple ordinal ordering of response options by their ranks, starting with 1 for the least happy response, followed by linear stretch to scale 0-10, 2) assigning fixed weights to particular response options, such as 8 for 'happy', 3) assessing weights in the context of the response scale, using native speakers as judges, and 4) deriving weights of verbal response options using a reference distribution obtained in the same year from answers to a similar survey question rated on a numerical response scale. These methods have been developed over the years for the World Database of Happiness and are described in DeJonge 2017.

4.1 Verbal response options weighted equally, by linear stretch to range 0-10 *(Linear Stretch Method)*

The Linear Stretch Method is a traditional method to transform means obtained using a verbal scale to a common range from for example 0 to 10. In the Linear Stretch Method response scales are stretched to obtain the common range. This is done in such a way that the number corresponding to the response option representing the worst case in the scale is always projected onto 0 and the number representing the best case in the scale onto the highest value of the range, with all the intermediate options given equally distanced numbers in between: for the 4-point verbal scale taken from the Japanese LIN survey the transformation to a 0-10 scale applying the Linear Stretch Method results in 0, 3.33, 6.66 and 10. A sample mean is calculated as the weighted average of the transformed ranks of the response options using the relative frequencies as weights. This gives a mean value between 0 and 10.

The disadvantages of assumed equidistance, the ignoring of the verbal labels of the response options and the assumption of a discrete distribution within the population still remain in the Linear Stretch Method.

4.2 Verbal response options weighted by fixed values and stretch to range 0-10 *(Thurstone Conversion)*

A method to cope with the disadvantages of assumed equidistance and the ignoring of the verbal labels of the response options is found in a method where a group of experts is employed to rate the verbal labels of response options on a common numerical scale. We will refer to this method as the Thurstone Conversion, as Jones and Thurstone (1955) were pioneers in applying this method.

Following their example, Veenhoven (1993) and 12 co-workers rated the degree of happiness denoted by the verbal labels of commonly used survey questions on a numerical 0 to 10 scale. The questions on life satisfaction from the Japanese LIN survey were also included in this rating exercise, for which the labels of the response options were translated literally from Japanese to English. For the LIN questions this resulted in the fixed values given in Table 1.

Table 1
Fixed values response options LIN questions

Rank response option	1958-1963	1964-1991	1992 to date
4	8.5	7.0	8.5
3	6.9	6.5	6.8
2	4.0	4.0	4.0
1	1.1	1.2	2.9

In the Thurstone Conversion, the numerical values resulting from the expert rating are assigned as fixed values to the corresponding original verbal response options of a given survey question to obtain a transformed mean. This transformed mean can be calculated as the weighted average of the fixed values on scale 0-10 using the relative frequencies measured in a wave of surveying as weights.

The Thurstone Conversion is still the common method to transform mean values to scale 0-10 in the WDH for verbal scales with at most 7 response options. This method is also applied in the WDH to obtain comparable average scores. To this day, these results are used to transform responses in the WDH with scales for which linear stretching falls short.

Although the Thurstone Conversion overcomes the disadvantages of assumed equidistance and ignoring of the labels that are associated with the Linear Stretch Method, an important weak point that remains is that the context of the response options is disregarded: 'Very happy' may, for example, be interpreted differently if it is followed by the response options 'Extremely happy' in a 7-point scale, than if it is the top anchor-point option in a 4-point scale. Furthermore, the Thurstone Conversion implicitly assumes a discrete distribution of happiness in a population, whereas it is more realistic to assume that this distribution is continuous.

A last disadvantage of Thurstone Conversion that we mention here, is that the assessment was done by one group of experts on labels of the response options which were translated literally from Japanese to English. It is unsure whether or not a different group of experts would give the same results or whether the translation from Japanese to English is an influencing factor.

4.3 Verbal response options assessed by native speakers on a numerical scale 0-10 in the context of the response scale

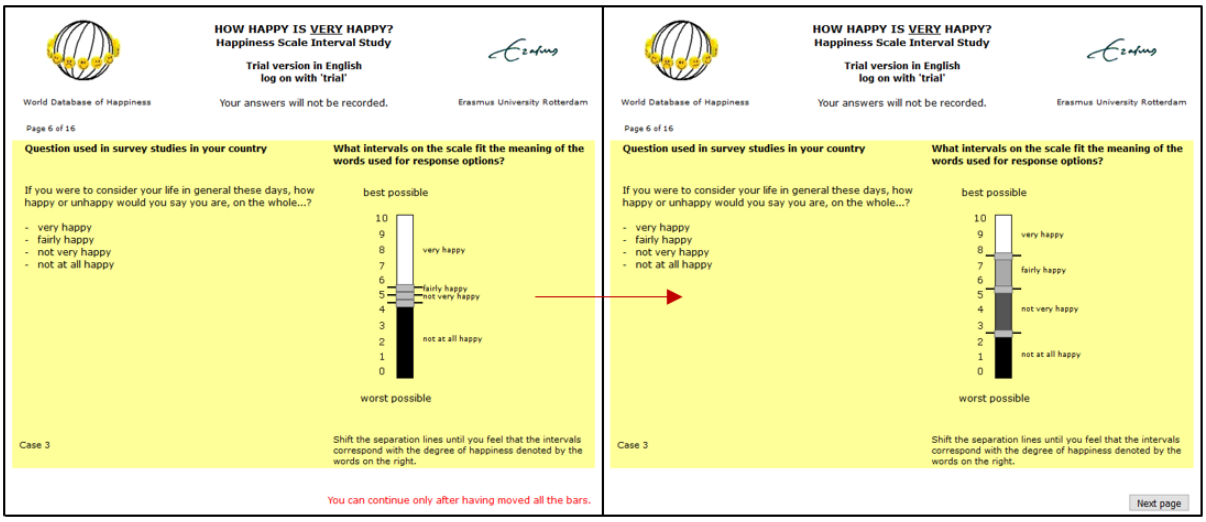
(Scale Interval Method)

The next method we discuss is the Scale Interval Method which was developed to tackle the shortcomings of the Thurstone Conversion (Veenhoven 2008). In the Scale Interval Method judges assess the points on a bounded continuum from 0 to 10, in which verbal response options for a given response scale transit from one to another. This can be done using the web-based Scale Interval Recorder (Veenhoven and Hermus 2006). Using this recorder, a series of survey questions is presented on a computer screen to judges. Questions are presented sequentially on the left side of the screen and each question presented consists of a leading question and its

corresponding verbal response scale with options given in the judges' mother tongue. An example of a question from the trial version[15] of the Scale Interval Recorder is given on the left of Figure 5. The judges have to shift the sliders on the vertical bar scale on the right side of the survey question until they feel that the intervals on the vertical bar correspond to the meaning of the words as used for the verbal response options. A possible result of this shifting of the sliders is given on the right of Figure 5.

The approach to scale transformation used in the Scale Interval Method differs essentially from that used in the Linear Stretch Method and Thurstone Conversion, as the response options in the primary scale are not considered to be discrete points, but to be intervals that each represent a part of the continuum from 0 to 10 where the value of the latent[16] variable can be found. The assessments by all judges are averaged for each transition point between two consecutive response options, to define the boundaries of these intervals. The lower boundary of the response option labelled by the worst option is by definition always equal to 0. The upper boundary of the response option labelled by the best option is by definition always equal to 10.

Figure 5
Example Scale Interval Recorder, before and after assessment



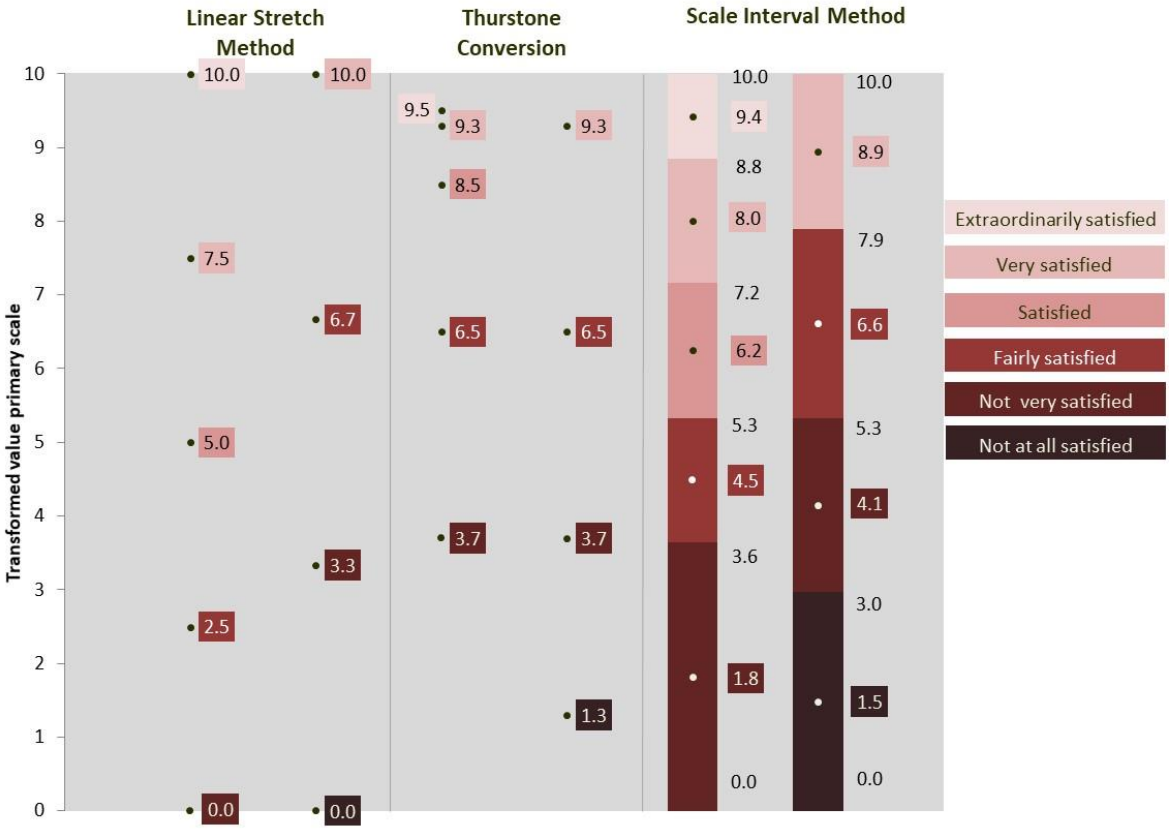
In the Scale Interval Method, it is assumed that the distribution of the latent variable for happiness can well be approximated by a beta distribution (Kalmijn 2010; Kalmijn, Arends and Veenhoven 2011), because 1) a beta distribution has a two-side bounded domain, 2) beta distributions cover a wide class of distributions shapes based on two shape parameters and 3) a beta distribution is independent of the format of the primary, discrete scale used for surveying. The mean μ of a beta distribution with shape parameters α and ß on the continuum from 0 to 10 is equal to $10 * \frac{\alpha}{(\alpha+\beta)}$. For a more elaborate description of this Continuum Approach we refer to

[15] http://www.risbo.org/fsw/english-trial/
[16] We use the term latent variable for unobservable variables such happiness and life satisfaction, which can take any value on a continuum and are normally measured indirectly through observed scores often obtained using a discrete scale.

Kalmijn (2010) and DeJonge et al (2017, Ch. 7). An estimated population mean in the Scale Interval Method for a given survey question and wave of surveying can be calculated on basis of the shape parameters of the beta distribution that best fits[17] the measured cumulative frequency distribution and the upper boundaries of intervals corresponding to the response options obtained from the assessment.

To get a more illustrative impression of how the Scale Interval Method related to the Linear Stretch Method and the Thurstone Method, we have given an example of the application of each method to a 4- and a 5-point scale for measuring life satisfaction in Figure 6 (see also Figure 2.2, DeJonge et al 2017, p. 23).

Figure 6
Comparison of the three scale homogenization methods discussed



Looking at Figure 6, it can be seen that when using the Linear Stretch Method and in Thurstone Conversion, after the transformation there is still a discrete number of response options. For the Scale Interval Method it is probably better to speak of a conversion of the scale instead of a transformation. In the Scale Interval Method the discrete points of the primary scale are converted to a series of connected intervals that together span the entire continuum from 0 to 10. Moreover, in the Scale Interval Method, the context of the scale is taken into account. Looking at Figure 6 it is clear that the response option 'Satisfied' represents an interval from 5.3 to 7.2 in the context of the 5-point scale and from 7.9 to 10.0 in the 4-point scale as a result of the application of the Scale Interval Method.

[17] The parameters α and ß are estimated using the maximum likelihood method.

Although with the introduction of the Scale Interval Method a step forward has been made, one from discrete to continuous, the method has a drawback that we have not previously mentioned. This drawback is that the method does not take into account that the construction of the scale may be such that the options respondents can choose from and the position of these options in the scale are not in accordance with their expectations. If, for example, the option 'Satisfied' is positioned in the middle of the scale, a respondent who is satisfied may be tempted to tick the next best option because that position on the scale is more in line with their perception of how they feel. Another example is when respondents have to choose between two options that are suboptimal for them such as when a satisfied respondent has to choose between 'Fairly satisfied' and 'Very satisfied'. Apparently a majority of the satisfied respondents tend to prefer the option 'Very satisfied' over the option 'Fairly satisfied', which will push the beta distribution to be more skewed to the right. The consequence is that it may occur that, after applying the Scale Interval Method, the converted means for two survey question used for measuring happiness in the same year in a representative sample of the population are not the same (DeJonge, Veenhoven and Arends 2014, p. 287; DeJonge et al 2017, 78-80).

4.4 Weights of verbal response options derived from the distribution of the response on a similar question in the same year using a numerical response scale

(Reference Distribution Method)

The last method we will discuss is the Reference Distribution Method (DeJonge et al 2017, Ch. 8). The first assumption for this method is that for a given year and a given population, the means after transformation for similar questions from different surveys will be equivalent. Secondly it is assumed that the latent variable is continuously distributed in the population and thirdly it is assumed that the response options of verbal scales are not equidistant. The first assumption is a response to the additional drawback we mentioned in relation to the Scale Interval Method.

In the Reference Distribution Method, a reference distribution is used to make survey results from different surveys or those measured with the same survey but with different survey questions in different periods of time comparable. To explain what a reference distribution is, we make use of the question on life satisfaction taken from the World Values Survey (WVS). This question consists of the leading question "All things considered, how satisfied are you with your life as-a-whole these days?" and 10 numerical response options. The cumulative frequency distribution for this question measured in Japan in 2005, as vertical bars equidistantly distributed[18] over the 0 to 10 continuum is shown in Figure 7.

The curve shown in Figure 7 is the beta distribution that according to the Continuum Approach best fits the boundaries and cumulative frequencies distribution of the WVS question in 2005 for Japan. The shape parameters of this best fit beta

[18] For numerical scales we choose to fix the upper boundaries of the response options on the 0 to 10 continuum equidistantly (Kalmijn 2013).

distribution are equal to α = 3.98 and ß = 2.20 and give an estimated population of 6.45.

The best fit beta distribution to the WVS question can now be used as a reference distribution for deriving the boundaries between response options for other survey questions used for measuring life satisfaction in Japan in 2005. This is illustrated in Figure 8 for the question from the LIN survey used since 1992 in the translation in English and the measured frequency distribution in 2005 in Figure 8. The cumulative frequency distribution for the LIN question in 2005 is given as a stacked bar on the left of Fig 8. The boundaries between the response options are equal to the value on the 0-10 continuum where the cumulative frequency of the LIN question is equal to that of the reference distribution.

The boundaries for the LIN question from the reference distribution are equal to 4.04 for the option 'Dissatisfied', 6.03 for the option 'Somewhat satisfied', 8.84 for the option 'Fairly satisfied' and by default 10 for the option 'Satisfied'. The converted population mean for the LIN question in 2005 is the same as that for the reference distribution which we found to be equal to 6.45. The reference boundaries of the response options of the LIN question found in this way are kept fixed and are used to estimate the beta distributions that best fit the cumulative frequencies of other waves in which the same question has been employed. The estimated mean on the 0 to 10 continuum for each of these waves is equal to the mean of the corresponding best fit beta distribution.

Figure 7
Application of the Continuum Approach to derive a reference distribution

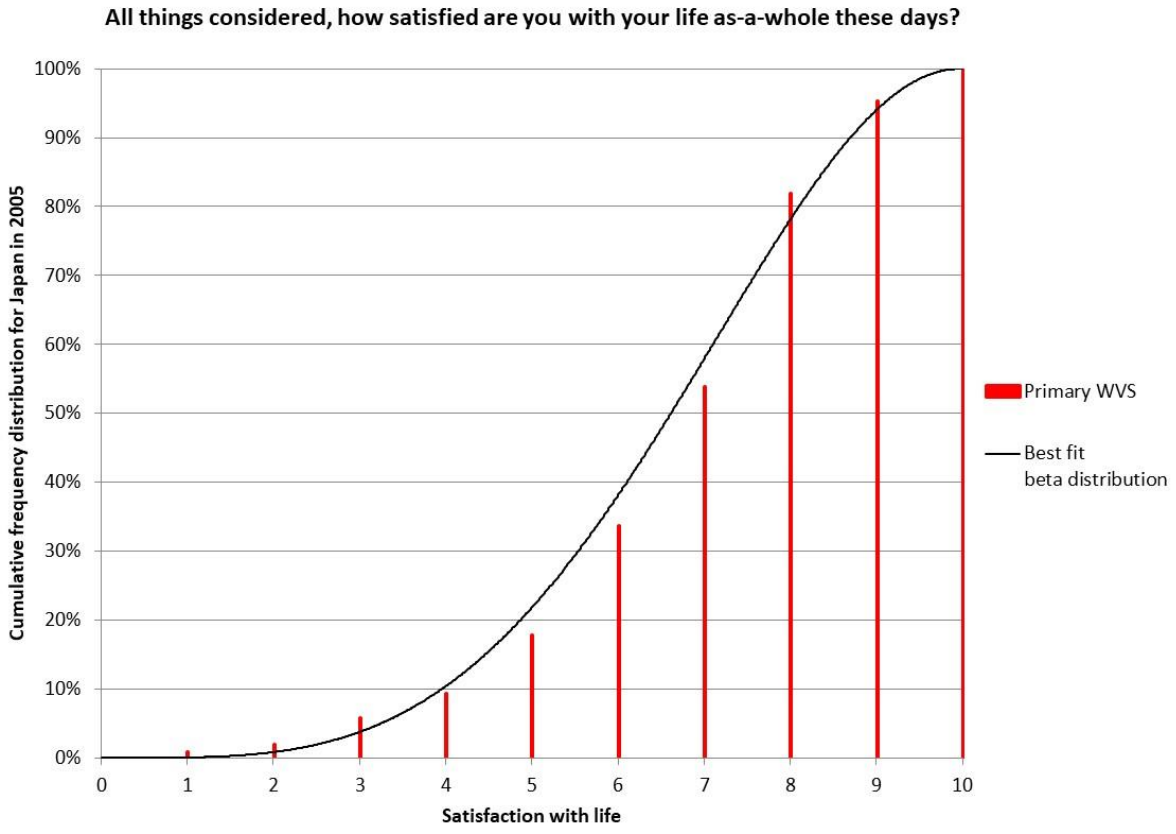
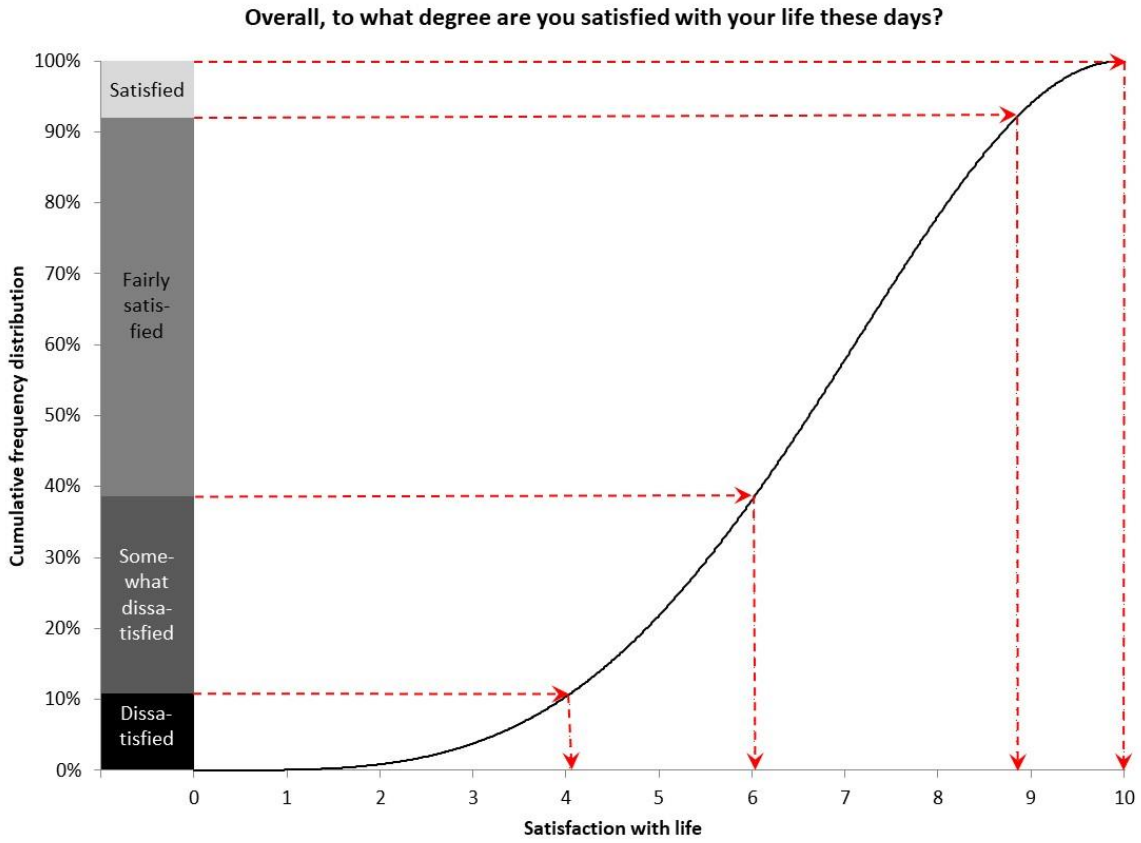


Figure 8
Illustration of the Reference Distribution Method



In long-term survey programs the survey question or the mode of surveying, such as face-to-face interviewing or paper-and-pencil surveying do not always remain constant for long lasting time series and may be changed. If so, it is plausible that such changes will influence the position of the boundaries between the response options. In such a situation, this position has to be reconsidered and presumably determined anew. To derive new boundaries that comply with the new survey mode or changed survey question, the original reference distribution should not be used. Instead the best fit beta distribution, given the boundaries derived from the original reference distribution and the frequency distribution of the survey results in the year prior or equal to that in which the mode or the question was changed, should be selected as a new reference distribution. Whether the new reference distribution should be based on the survey results for the year the change occurred or for the year prior to the change(s), depends on whether there have been two surveys in the year of the change(s): in the ideal situation a survey will be fielded in two versions to get insight into the effect of any change(s). In this case the new reference distribution can be based on the survey results for the same year as the change. If unfortunately no double measurements are available, but the survey results show minor changes from year to year, the best fit beta distribution estimated for the year prior to the year the questionnaire was changed can be used as a proxy. This is, however, a suboptimal solution the results of which should be interpreted with caution.

5 APPLICATION OF THESE 4 SCALE HOMOGENIZATION METHODS ON THE HETEROGENOUS TIME-SERIES DATA ON HAPPINESS IN JAPAN

We will now inspect what the transformed trends in happiness in Japan are for each of the scale homogenization methods we described, when applied to the available survey data on happiness in Japan (cf. Section 1.2). In each figure that we present, we have included the trend in happiness in Japan based on the Rank Method for reasons of comparison. Although the data we had available span the time period from 1958 until 2013, we restricted the results to the period from 1958 to 2007, to comply with the period Suzuki and Stevenson & Wolfers studied (cf. Sections 3.2 and 3.3). In addition and in contrast to what the above authors did, we only used one wave of data per year, this did not have a noteworthy effect on the results.

5.1 Trend when equal weights are assigned to all response options *(Linear Stretch Method)*

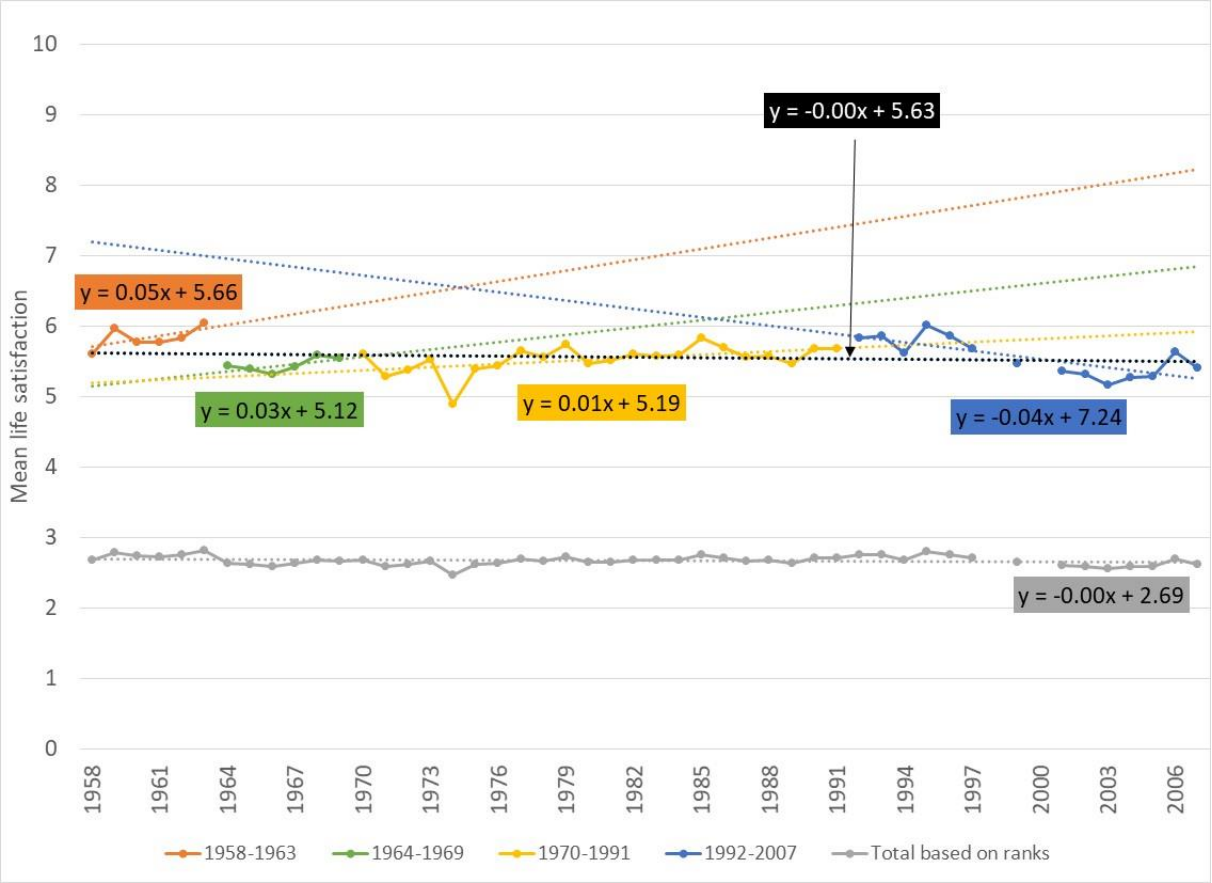
Recall from Section 4.1 that when applying the Linear Stretch Method for the 4-point verbal scale taken from the Japanese LIN survey a transformed sample mean on scale 0 to 10 was calculated as the weighted average of the transformed values given to the response options of 0, 3.33, 6.66 and 10 using the relative frequencies as weights. The transformed time series based on the Linear Stretch Method are given in Figure 9. The time series is split into four time periods in Figure 9, based on the years in which the survey question was changed. The linear trend is given for each of these time periods and for the entire period 1958-2007. The time series of the unconverted means, based on the Rank Method, is given in grey for the entire period as is the linear trend for this period.

Like Stevenson and Wolfers, we find that application of the Linear Stretch Method, gives an upward trend in life satisfaction in the first three sub-periods and a downward trend in the last period. The slope of the upwards trends is smaller the later the period in time. When we combined the homogenized means for the sub-periods into one time series, we were left with a transformed time series for mean life satisfaction from 1958-2007 with values between 4.90 and 6.05 and a trend line with slope 0. We concluded that life satisfaction in Japan did not on average change between 1958-2007 on basis of the Linear Stretch Method.

5.2 Trend when fixed weights are applied to specific verbal response options *(Thurstone Conversion)*

Recall from Section 4.2 that the values given to the response options of the LIN question used for the calculation of a converted sample are not equidistant as in the Linear Stretch Method but instead equal to 1.1, 4.0, 6.9, and 8.5 for the period 1958-1963, 1.2, 4.0, 6.5 and 7.0 for the period 1964-1991 and 2.9, 4.0, 6.8 and 8.5 for the period 1992 onwards. What the effect of these values is on the trends in mean happiness is shown in Figure 10.

Figure 9
Transformed time series of mean happiness in Japan, Linear Stretch Method



The results shown in Figure 10 are similar to the results shown in Figure 9. The mean values and the magnitude of the fluctuations seen in Figure 10 are in line with the means and fluctuations of the transformed time series based on the Linear Stretch Method. Once again, we concluded that life satisfaction in Japan has not, on average, changed between 1958-2007.

5.3 Trend when verbal response options are homogenized using the Scale Interval Method

The boundaries between the response options of questions used for measuring happiness in Japan, were assessed in the Scale Interval Study covered by the studies Japanese-6 and Japanese-7[19]. The assessments were done by almost 200 students mainly from the School of Business administration of the Aoyama Gakuin University in Tokyo. Three of the four questions from the Japanese LIN survey used for measuring life satisfaction in Japan since 1958 were included in the study, for the question used from 1964-1969 we applied the boundaries obtained for the period 1970-1991, analogous to how the entire period from 1964-1991 was treated for the Thurstone Conversion. The results of the assessment are given in Table 2.

[19] See the study list at https://worlddatabaseofhappiness.eur.nl/scalestudy/scale_fp.htm.

Figure 10
Converted time series of mean happiness in Japan, Thurstone Conversion

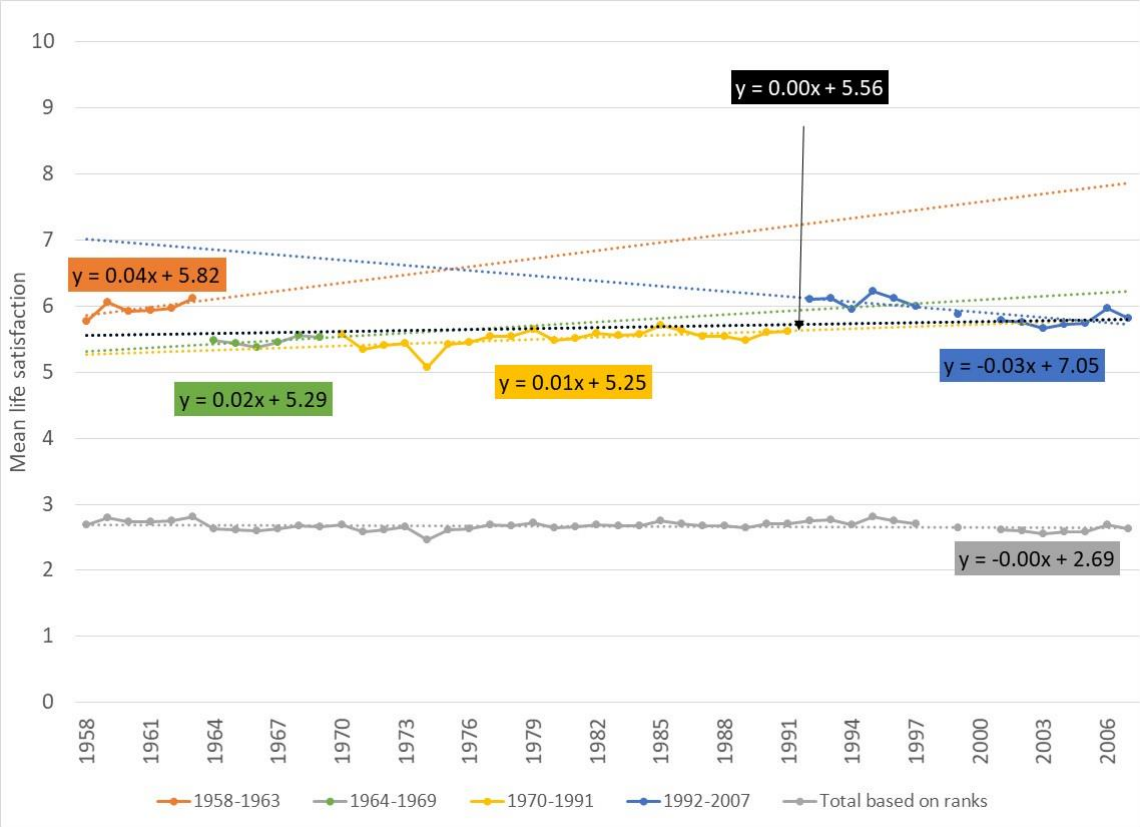


Table 2
Assessed boundaries between response options LIN questions

Rank response option	1958-1963	1964-1991	1992 to date
4	10.00	10.00	10.00
3	7.80	7.81	7.59
2	4.53	4.63	4.43
1	1.81	1.81	1.86

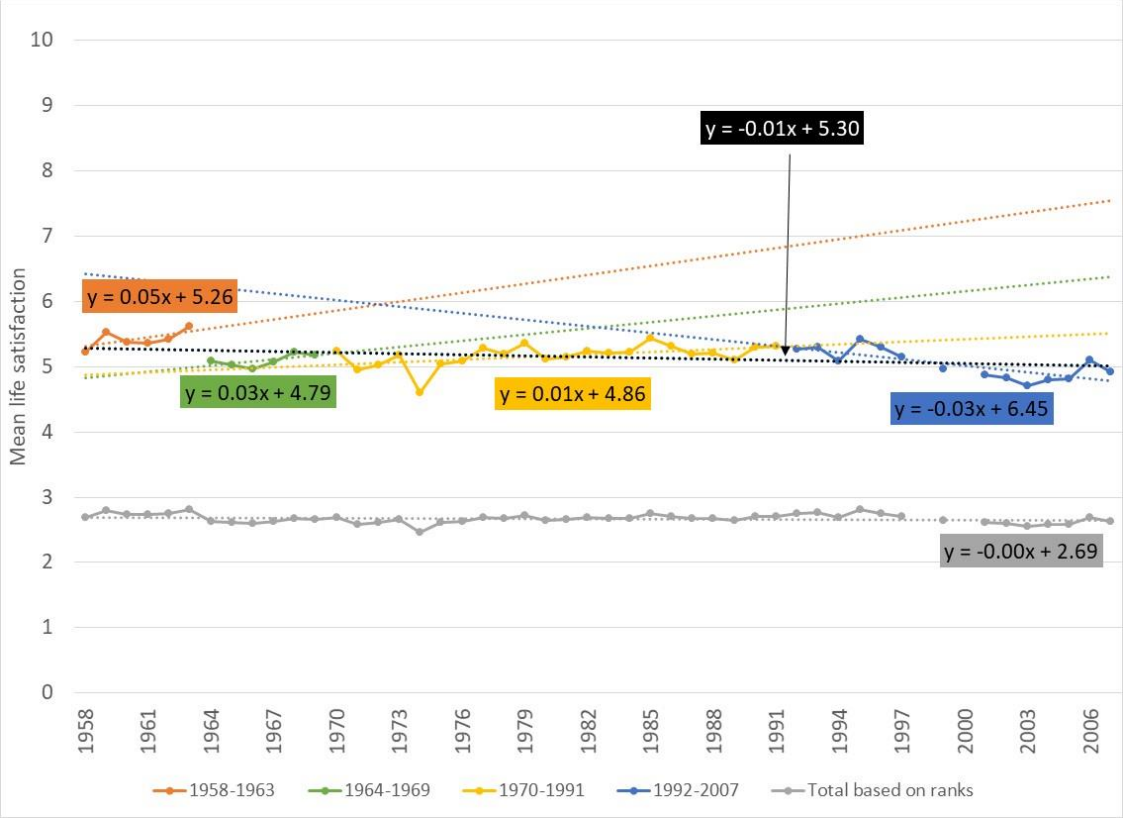
As can be seen from Table 2, there are small differences between the boundaries assessed for each period of time.

Given the intervals representing the response options of the LIN question in each period, we estimated the parameters α(t,p) and ß(t,p) of the best fit beta distribution for each year t in period p. We used these parameters to estimate the mean happiness in the population in each year in the period 1958-2007 as $10 * \frac{\alpha(t,p)}{(\alpha(t,p)+\beta(t,p))}$. The result of these estimates is given in Figure 11.

The converted means based on the application of the Scale Interval Method fluctuate between a somewhat lower range than when applying Linear Stretch or Thurstone Conversion to the LIN question and had a minimum value 4.61 and maximum value of 5.62. On basis of the Scale Interval Method we once again had to

conclude that life satisfaction in Japan had not, on average, substantially changed between 1958-2007.

Figure 11
Converted time series of mean happiness in Japan, Scale Interval Method



5.4 Trend when verbal response options are homogenized using the Reference Distribution Method

The boundaries between the response-options of questions used for measuring happiness in Japan are derived from a reference distribution when the Reference Distribution Method is applied. In Section 4.4 we explained the Reference Distribution Method, using the 2005 wave of the WVS to provide a reference distribution. In this section, however, we will first show the scale homogenization results for the LIN question using this method, when the best fit beta distribution to the frequency distribution measured with the LIN question in 1992 and the boundaries assessed for this question in the Scale Interval Study is used as an initial reference distribution. Effectively we begin by combining the Scale Interval Method and the Reference Distribution Method.

The converted means for the period 1992 to 2007 in Japan were calculated on the basis of the parameters of the best fit beta distributions to the boundaries between the response options assessed in the Scale Interval Method and the frequency distribution measured in each wave. Next, we used the best fit beta distribution thus found for the 1992 Japanese wave, as a reference distribution to derive the boundaries between the response options of the question used between 1970 and 1991. We fix these boundaries to the values where the cumulative

frequency distribution of the 1991 wave measured with the LIN question was equal to the reference distribution. Note: this is a suboptimal solution, since there was no double measurement done in 1992, i.e. one with the question used from 1970 and 1991 and one with the question used from 1992 onwards. The estimated population for 1991 is thus equal to the estimated population mean in 1992.

Using the fixed boundaries derived for the question used in the Japanese LIN survey in the period 1970 to 1991, we estimated best fit beta distributions for all the other waves in this period. The best fit beta distribution found for the 1970 wave then served as a reference distribution to the derive the boundaries between the response options for the question employed from 1964 to 1969 and similarly, the best fit beta distribution found for the 1964 wave then served as a reference distribution to the derive the boundaries between the response options for the question employed from 1958 to 1963. The converted time series for mean happiness in Japan following from this procedure is given in Figure 12.

The converted means, shown in Figure 12, fluctuated between 4.53 and 5.43 when applying the Reference Distribution Method in combination with the Scale Interval Method. The trend for the period 1992-2007 in Japan was identical to the trend for this period shown in Figure 11, since for this period we applied the Scale Interval Method. The Reference Distribution Method was helpful for obtaining converted means for the period 1964-1969. The trend of converted means for the entire period from 1958-2007 had a zero slope. This again led to the conclusion that life satisfaction in Japan did not on average change between 1958 and 2007.

The converted time series for Japan so far had values that were much lower on the continuum from 0 to 10, than the mean value of 6.45 we found for the beta distribution that best fits the 2005 wave of the WVS and ten equidistant boundaries for the ten numerical response options. Therefore we also applied the Reference Distribution Method using the 2005 wave of the WVS to derive an initial reference distribution.

We used the reference distribution to derive the boundaries between the response options of the Japanese LIN question used between 1992 and 2007. As described in Section 4.4, these boundaries were equal to 4.04 for the option ‘Dissatisfied’, 6.03 for the option ‘Somewhat satisfied’, 8.84 for the option ‘Fairly satisfied’ and by default 10 for the option ‘Satisfied’.

Using the above described combination of the Reference Distribution Method and the Scale Interval Method, we estimated reference distributions for the questions used in the three other periods of the Japanese LIN survey. The boundaries found for all periods using the Reference Distribution Method are given in Table 3. The boundaries slightly differ per period. The converted time series for mean happiness in Japan following from this procedure is given in Figure 13.

Figure 12
Converted time series of mean happiness in Japan, Reference Distribution Method, initial reference distribution based on 1992 wave LIN question and boundaries based on the Scale Interval Method

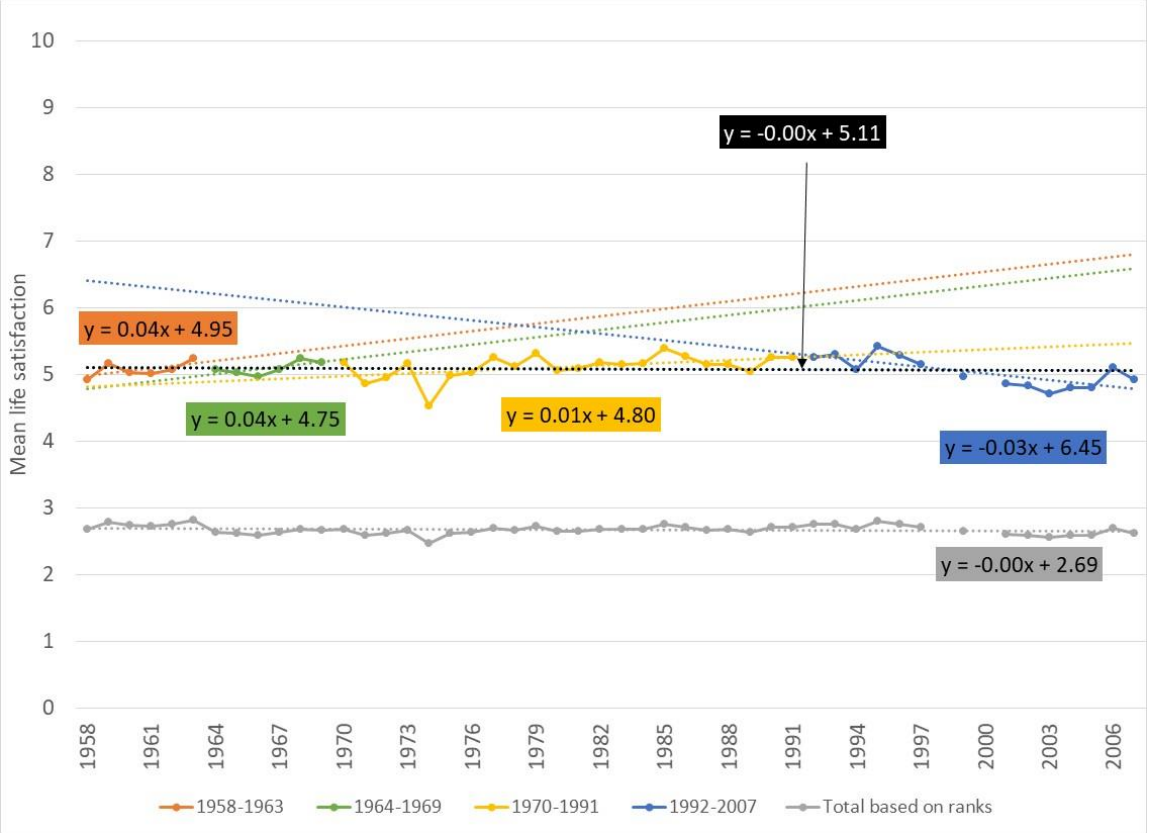
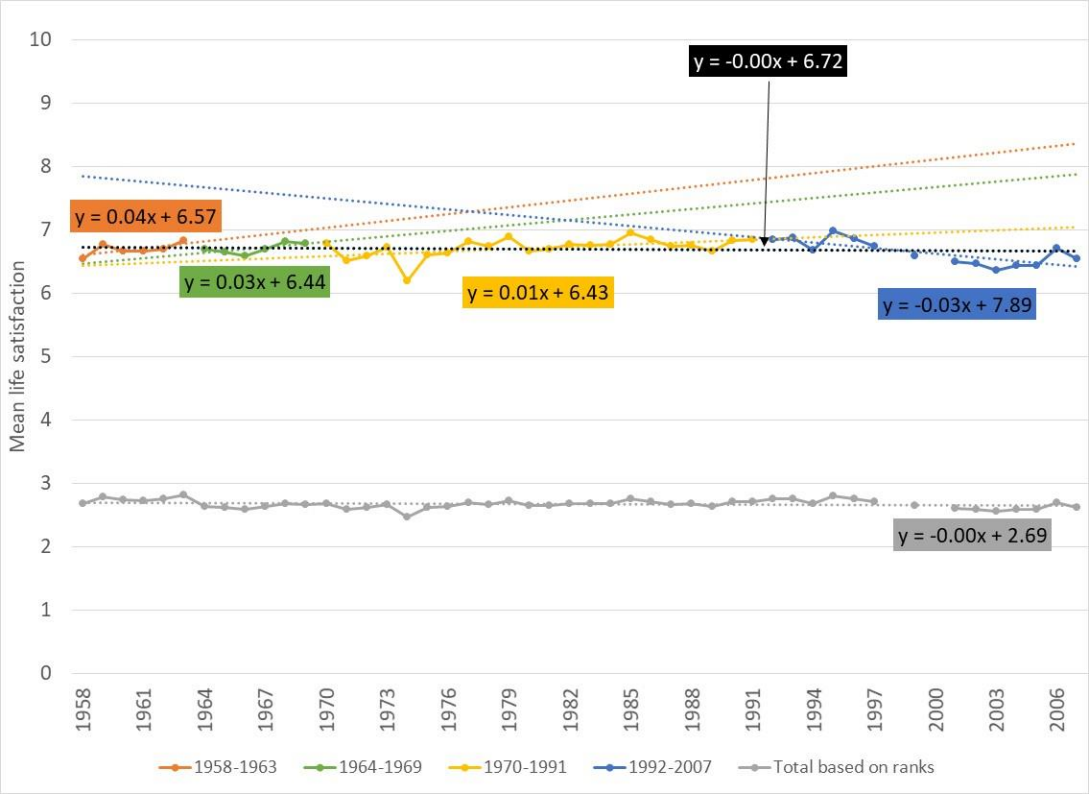


Table 3
Boundaries between response options LIN questions based on the Reference Distribution Method, starting with the 2005 wave of the WVS

Rank response option	1958-1963	1964-1969	1970-1991	1992 to date
4	10.00	10.00	10.00	10.00
3	8.30	9.07	9.05	8.84
2	6.23	6.28	6.18	6.03
1	3.94	3.67	3.64	4.04

The converted means shown in Figure 13 fluctuated between 6.20 and 6.99 when applying the Reference Distribution Method using the 2005 wave of the WVS as a seed. These values are more in line with measurements using numerical scales for measuring happiness in developed countries. Similar trends as for the other scale homogenization methods were found, with a zero slope of the trend in the converted time series for the entire period from 1958-2007. This again led us to the conclusion that life satisfaction in Japan had not, on average, changed between 1958 and 2007.

Figure 13
Converted time series of mean happiness in Japan, Reference Distribution Method, initial reference distribution based on the 2005 wave of the World Values Survey



In presenting these results, we only made use of data taken from the LIN and the WVS. We mentioned four other surveys that have been used to measure happiness in Japan in Section 1.3. In the appendix we give the results of an application of the Reference Distribution Method making use of the data from these other surveys to derive reference distributions from which to convert the time series on happiness taken from the LIN survey.

6 DISCUSSION

6.1 Fit with trends in other items

We applied four scale homogenization methods to bring the time series of mean life happiness in Japan measured using a verbal scale to a numerical level on the 0-10 continuum. We accounted for the assumption that the latent variable is continuously distributed within a population[20] and for the changes in the survey questions used over time only in the Scale Interval Method and the Reference Distribution Method. What the different methods have in common, however, is that from the converted time series we had to conclude that happiness has not changed in Japan between 1958 and 2007.

6.2 Possible reasons for stagnating happiness in Japan

In the methods we have applied, we first homogenized the data measured using different survey questions and then looked at the trend in mean happiness in Japan. We believe that it is justified to conclude that happiness has stagnated in Japan, despite the enormous economic growth the country has gone through. In this section we will give some possible reasons for this stagnation in mean happiness.

Adaptation to economic growth

A first possible reason is adaptation to the new level of affluence and a rise in aspirations for material comforts. The individual's reported level of happiness will reflect the gap between what they actually receive and what they expected to receive (Clark, Kamesaka and Tamura 2015, p.527; Clark, Frijters and Shield 2008, p.34). A related process is social comparison, if happiness depends on the difference in wealth with the people around you, you will not get any happier if everybody gets richer in the country. These explanations are at the core of the 'Easterlin Paradox', which holds that economic growth does not make us any happier since "A rising tide lifts all boats'' (Easterlin 1995). Yet the Easterlin Paradox describes exceptionally cases rather than the rule, during the last 50 years economic growth has been accompanied by rising happiness in most countries (Veenhoven & Vergunst 2014). Japan and the USA are among the few cases in which happiness has remained at the same level in spite of economic growth. So the question is: Why have the Japanese profited less from economic growth than people in most other countries?

Costs of economic growth

A possible answer to the above question is that the costs of economic growth have been higher in Japan than in other countries. The Japanese have a reputation for a high work-ethic and long working hours coming with extremes which manifest in two typically Japanese phenomena.

The first is *karoshi*, which according to the Japan Institute for Labour Policy and Training (JILPT) is defined as death caused by cerebrovascular disease or ischemic heart disease resulting from excessive work (JILTP 2014, p. 104) or in

[20] Unlike in the method of ordered probit it is not assumed that the latent variable is normally distributed.

short the occurrence of death by overworking (Kamesaka and Tamura 2017). The second is *karojisatsu* which is defined as suicide induced by overwork and work-related depression (JILPT, p. 104). These terms are associated to very long working hours experienced by many workers in Japan, and these may have an effect on how happy people feel.

Besides the long working hours in the last decades, the Japanese labor market has changed enormously (JILPT 2014). The number of regular staff workers has dropped from 82% in 1988 to 65% in 2012, where especially the younger workers are confronted with non-regular work. There is also a growing gap between the wages earned by regular staff workers and non-regularly employed workers. This may come with feeling less secure about the future among the last group of workers, which might have a negative impact on subjective well-being. Japan has also made a strong transition in its employment structure. A striking example is that the share of people working in agriculture and forestry dropped from 38% in 1953 to less than 4% in 2010. In contrast the share of professional and technical workers grew from 4.4% of all workers to 15.8% in that same period (JILPT 2014, pp. 37-39). As long as this transition continues asking worker to re-skill, there may be a stagnating effect on mean happiness.

Other aspects which may hamper an increase in mean happiness in Japan that can be found in the JILPT report, are the drop in the fertility rate since 1958 which has led to a rapidly ageing society, a lower number of people getting married and a larger share of women participating on the labor market combined with changes in the work-life balance, relations between men and women and the perspective Japanese people have on society.

The above is not intended to be an exhaustive list of possible reasons for stagnating happiness in Japan. We mainly want to stress that there is more than just macro-economic indicators that need to be considered when explaining trends in happiness in Japan. Likewise, when assessing the effects of economic growth on happiness, one should keep an open mind for parallel developments that may veil positive or negative correlations.

7 CONCLUSIONS

When happiness is measured using different survey questions, before drawing any conclusions the results should be homogenized before a trend can be identified. The Scale Interval Method and the Reference Distribution Method are suitable for this.

Using these methods, we find that the Japanese have not got any happier over the last 50 years. If the economy has had a positive effect on mean happiness at all, this gain has been counterbalanced in some way, possibly by the costs of economic development, such as the long working hours or exogenous factors such as the pain(s) of cultural change.

Acknowledgements We would like to thank Mr. Susumu Kuwahara[21] and Mr. Ryoichi Watanabe of the Cabinet Office, Government of Japan, for their comments and support. This work was supported by the Japanese Society for the Promotion of Science (JSPS) KAKENHI Grant Number 16H03640. The authors also thank Dr. Tije Euvermans, methodologist at the Erasmus University Rotterdam, for his comments and Miranda Aldham-Breary MSc, senior volunteer at the WDH, for the English text editing.

[21] Mr. Susumu Kuwahara has moved to the Ministry of Foreign Affairs of Japan in August 2018.

REFERENCES

BBC (2006). *The happiness formula: Opinion poll GfK-NOP poll 421059.* Report available at: http://news.bbc.co.uk/2/hi/programmes/happiness_formula/4771908.stm.

Clark, A.C., Frijters, P. & Shields, M.A. (2007). Relative income, happiness and utility: an explanation for the Easterlin paradox and other puzzles. *IZA Discussion Papers, No. 2840,* Institute for the Study of Labor (IZA), Bonn. Available at: https://www.econstor.eu/handle/10419/34701.

Clark, A.C., Kamesaka, A. & Tamura, T. (2015). Rising aspirations dampen satisfaction. *Education Economics*, 23(5), 515–531. Available at: http://dx.doi.org/10.1080/09645292.2015.1042960.

Cummins, R.A. (2003). Normative life satisfaction: measurement issues and homeostatic model. *Social Indicators Research 64*, 225-240.

Dalai Lama, & Cutler, H (1998). *The art of happiness: A handbook for living.* Riverhead Press, USA.

Daykin, A.R., & Moffat, P.G. (2002). Analyzing Ordered Responses: A Review of the Ordered Probit Model. *Understanding Statistics*, 1(3), 157-166. Lawrence Erlbaum Associates.

DeJonge, T., Veenhoven, R. & Arends, L.R. (2014). Homogenizing responses to different survey questions on the same topic. Proposal of a Scale Homogenization Method using a Reference Distribution. *Social Indicators Research*, 117(1), 275-300. doi: 10.1007-s11205-013-0335-6.

DeJonge, T. Kalmijn, W. & Veenhoven, R. (2017). *Diversity in survey questions on the same topic: Techniques for improving comparability*. Social Indicators Research Book Series. Volume 68, ISBN: 978-3-319-53260-8 (Print) 978-3-319-53261-5.

Diener, E. & Diener, C. (1996). Most people are happy. *Psychological Science 7*, 181-185.

Diener, E., & Oishi, S. (2004). *Are Scandinavians happier than Asians? Issues in comparing nations on subjective wellbeing.* In: Columbus, F.; Ed.: "Asian Economic and Political Issues", 10, 1–25.

Easterlin, R.A. (1974). *Does Economic Growth Improve the Human Lot? Some Empirical Evidence.* In: David, P.A., Melvin, W.R.; Eds..: "Nations and Households in Economic Growth", Academic Press, 1974, New York, USA, 89-125.

Easterlin, R.A. (1995). Will Raising the Incomes of All Increase the Happiness of All?

Journal of Economic Behavior and Organization, Vol. 27, 35–47.

Frijters, P., Johnston, D.W., & Shields, M.A. (2008). Happiness Dynamics with Quarterly Life Event Data. *IZA Discussion Paper, No. 3604*.

Jones, L. V., & Thurstone, L. L. (1955). The psychophysics of semantics. An experimental investigation. *The Journal of Applied Psychology*, 39(1), 31–36.

Kalmijn, W.M. (2010). *Quantification of happiness inequality.* Dissertation, Erasmus University Rotterdam, The Netherlands. Enschede: Ipskamp Drukkers. Available at http://repub.eur.nl/pub/21777.

Kalmijn, W.M. (2013). From Discrete 1 to 10 Towards Continuous 0 to 10: The Continuum Approach To Estimating the Distribution of Happiness in a Nation. *Social Indicators Research*, 110(2), 549-557. Doi: 10.1007-s11205-011-9943-1.

Kamesaka, A. & Tamura, T. (2017). Work Hours and Anxiety toward Karoshi, *ESRI Discussion Paper Series*, 325. Economic and Social Research Institute, Cabinet Office, Tokyo, Japan. Available at: http://www.esri.go.jp/jp/archive/e_dis/e_dis325/e_dis325-e.pdf.

Stevenson, B. & Wolfers, J. (2008). Economic Growth and Subjective Well-being: Reassessing the Easterlin Paradox. *Brooking Papers on Economic Activity*, Washington DC, USA.

Suzuki, K. (2009). Are They Frigid to the Economic Development? Reconsideration of the Economic Effect on Subjective Well-Being in Japan. *Social Indicators Research*, 92, 81–89.

The Japan Institute for Labour Policy and Training (2014). *Labor Situation in Japan and Its Analysis: General Overview 2013/2014*. Tokyo, Japan. Available at: http://www.jil.go.jp/english/lsj/general/2013-2014/all.pdf.

Veenhoven, R. (1984). *Databook of happiness*. Dordrecht: Reidel Publishing. ISBN 90-277-1793-1.

Veenhoven, R. (1993). *Happiness in nations, subjective appreciation of life in 56 nations, 1946–1992 (Studies in social-cultural transformation, 2)*. Risbo, Erasmus University Rotterdam, Netherlands. Available at: https://personal.eur.nl/veenhoven/Pub1990s/93b-con.html.

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, Department of Social Sciences & Risbo Contract Research, The Netherlands. Available at: https://worlddatabaseofhappiness.eur.nl/scalestudy/scale_fp.htm.

Veenhoven, R. (2008). The International Scale Interval Study. (In V. Møller & D. Huschka (Eds.), *Quality of Life in the new millennium: 'Advances in quality-of-life studies, theory and research', Part 2: Refining concepts and measurement to assess cross-cultural quality-of-life* (pp. 45-58). Social Indicator Research Series, Vol. 35, Dordrecht: Springer Press.)

Veenhoven, R., & Berg, M. (2013). Has Modernization Gone too Far? Modernity and Happiness in 141 Contemporary Nations*. International Journal of Happiness and Development*, 1, 172– 95.

Veenhoven, R., & Vergunst, F. (2014). The Easterlin Illusion: Economic growth does go with Greater Happiness. *International Journal of Happiness and Development*, 1, 311-343.

Veenhoven, R. (2014). *Long-term change of happiness in nations: To times more rise than decline since the 1970s.* In: Sheldon, K. & Lucas, R. (Eds.), Stability of happiness: Theories and evidence on whether happiness can change', chapter 9, pp. 167-200, Elsevier Publishing.

Veenhoven, R. (2016). *Trend Average Happiness in Nations 1946-2015: How much people like the life they live.* World Database of Happiness, Trend Report Average Happiness, Erasmus University Rotterdam, 2016, Rotterdam, The Netherlands. Available at: https://worlddatabaseofhappiness.eur.nl/hap_nat/findingreports/TrendReport_AverageHappiness_1946-2015.pdf

Veenhoven, R. (2018a). *Subjective well-being in nations*. In Diener, E., Oishi, S. & Tay, L. (Eds.), Handbook of Well-being. Noba Scholar Handbook Series: Subjective well-being, DEF Publishers, Salt Lake City, Utah, USA. DOI: nobascholar.com.

Veenhoven, R. (2018b). *Happiness in Nations.* World Database of Happiness, Erasmus University Rotterdam, The Netherlands.

Veenhoven, R. (2018c). *Findings on Happiness in Japan.* World Database of Happiness, Erasmus University Rotterdam.

APPENDIX

COMPARISON OF TRENDS FOR DIFFERENT QUESTIONS TO DERIVE A REFERENCE DISTRIBUTION FROM

In the main part of this paper, in presenting results, we have only made use of data taken from the LIN and the WVS. We mentioned four other surveys that have been used for measuring happiness in Japan in Section 1.3. These other surveys are the

- Lifestyle Preference Survey (LPS), with a 5-point verbal scale question used to measure life satisfaction ten times between 1978 and 2002
- Future Life Survey (FLS), with a 5-point verbal scale question used to measure life satisfaction in 1995, 1997 and 1999
- East Asia Values Survey (EAVS), with a 5-point verbal scale question used to measure life satisfaction in 1988, 2002 and 2005
- Japanese National Character Survey (JNCS), with a 4-point verbal scale question used to measure life satisfaction in1993.

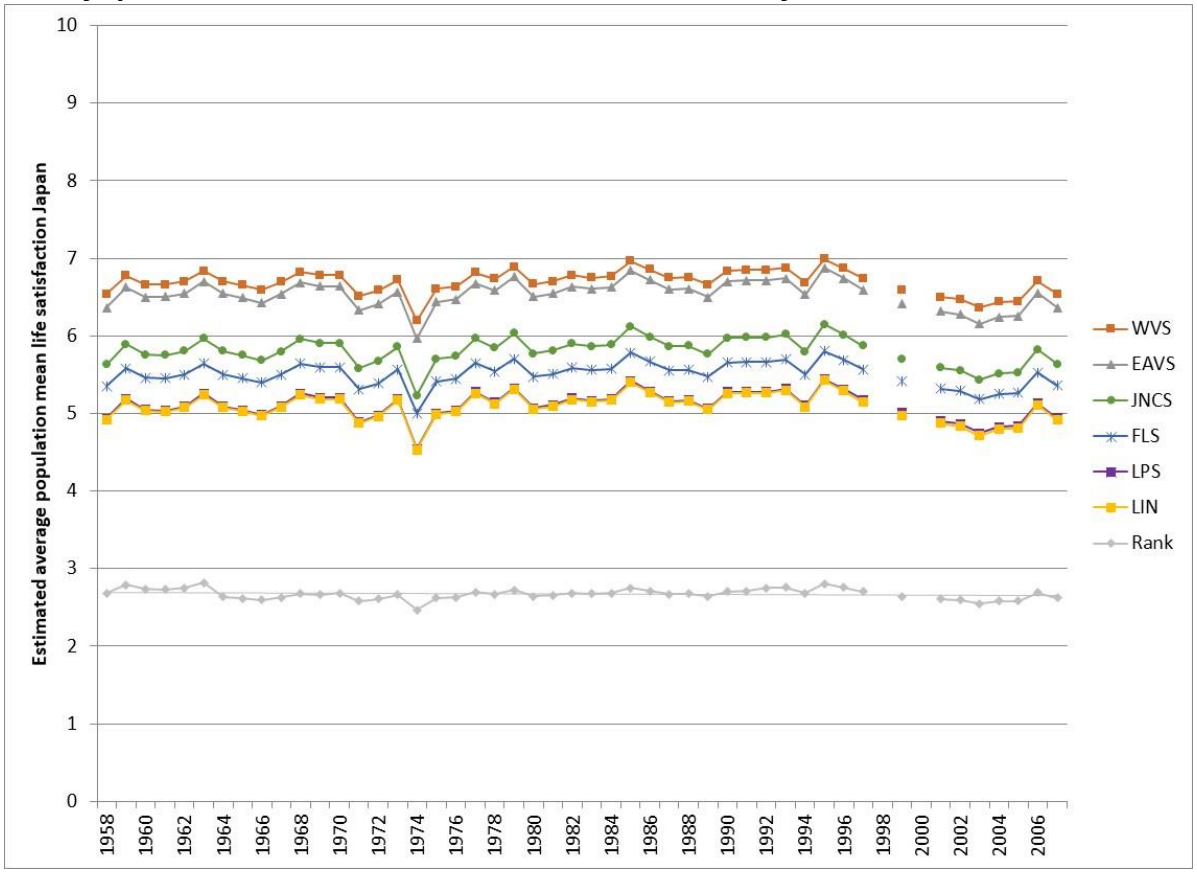
All these questions on life satisfaction can be found in the WDH and were also included in the Scale Interval Study referred to in Section 4.3 to be assessed by judges. The boundaries between the response options found for each of these questions in the Scale Interval Study can be used to derive a reference distribution for an appointed reference year. This reference distribution can then be used to convert the time series on life satisfaction from the Japanese LIN survey, if the reference year is chosen such that it is equal to a year in which the LIN survey was employed. The assessed boundaries between the response options of each question is given and the year appointed as the reference year for the survey are shown in Table 4.

Table 4
Boundaries between response options from the Scale Interval Study and appointed reference year

Rank response option	**LPS**	**FLS**	**EAVS**	**JNCS**
5	10.00	10.00	10.00	
4	7.89	7.91	7.64	10.00
3	5.60	5.46	5.42	7.65
2	3.64	3.57	3.51	4.67
1	1.54	1.58	1.75	2.02
Reference year	*2002*	*1999*	*2005*	*1993*

Using the reference boundaries shown in Table 4 and the frequency distribution of life satisfaction measured in the corresponding reference year, we estimated reference distributions to convert the time series from the LIN survey. The result of this conversion is given in Figure 14 in which we also included the conversion results starting with a reference distribution based on the LIN question or the WDH question presented in Figs. 12 and 13 and also the time series based on the Rank Method.

Figure 14
Converted time series of mean happiness in Japan, reference distributions based on survey question on life satisfaction from different surveys



It can be seen from Figure 14 that all converted time series develop in parallel but on a different level. Only the conversions based on the LIN question and the LPS are on the same level. The boundaries between the response options define the levels of the estimated means, but do not influence the evolution of the trends. The equations for the linear trend lines that can be draw through each converted time series in order of decreasing intercept are:

- WVS: 6.72 – 0.00x
- EAVS: 6.57 – 0.00x
- JNCS: 5.83 – 0.00x
- FLS: 5.53 – 0.00x
- LPS: 5.13 – 0.00x
- LIN: 5.11 – 0.00x

These results once more make us conclude that average happiness in Japan has not changed between 1958-2007.