

## Discussion Paper

# Measuring well-being

An analysis of different response scales

The views expressed in this paper are those of the author(s) and do not necessarily reflect the policies of Statistics Netherlands

2014 | 03

Jacqueline van Beuningen  
Karolijne van der Houwen  
Linda Moonen

## Explanation of symbols

.	Data not available
*	Provisional figure
**	Revised provisional figure (but not definite)
x	Publication prohibited (confidential figure)
–	Nil
–	(Between two figures) inclusive
0 (0.0)	Less than half of unit concerned
empty cell	Not applicable
2013–2014	2013 to 2014 inclusive
2013/2014	Average for 2013 to 2014 inclusive
2013/'14	Crop year, financial year, school year, etc., beginning in 2013 and ending in 2014
2011/'12–2013/'14	Crop year, financial year, etc., 2011/'12 to 2013/'14 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures..

### *Publisher*

Statistics Netherlands  
Henri Faasdreef 312, 2492 JP The Hague  
[www.cbs.nl](http://www.cbs.nl)

Prepress: Statistics Netherlands, Grafimedia  
Design: Edenspiekermann

### *Information*

Telephone +31 88 570 70 70, fax +31 70 337 59 94  
Via contact form: [www.cbs.nl/information](http://www.cbs.nl/information)

© Statistics Netherlands, The Hague/Heerlen 2014.  
Reproduction is permitted, provided Statistics Netherlands is quoted as the source.

# Measuring well-being

## An analysis of different response scales

Jacqueline van Beuningen, Karolijne van der Houwen and Linda Moonen

*This paper reports on three experiments relating to measuring well-being. Several scale types have been tested in three different experiments. First, we compared the current 5-point scales with verbal labels for happiness and satisfaction with life to a numerical 11-point scale ranging from 0 to 10 with verbal labels only for the end points of the scale. In a second experiment we tested three different types of numerical scales and in the third experiment we focused on respondents' interpretation of numerical scales.*

*Based on the results of these experiments Statistics Netherlands has decided to opt for a numerical 10-point scale ranging from 1 to 10 with verbal labels at the end points of the scale to measure well-being in the future. The results of the experiments have shown that these can be compared to international studies.*

**Keywords:** response scales, well-being

## 1 Introduction

Policy makers and researchers have become more and more aware of the fact that the well-being of people is as relevant – or perhaps even more relevant – than economic progress. This awareness was given a boost by the Beyond GDP initiative and by the Stiglitz, Sen and Fitoussi report (2009) on this subject. As a result international interest in well-being research has increased in recent years. Much debate within the research community has focused on how well-being should be defined, and measured.

Statistics Netherlands (SN) has measured subjective well-being (i.e. happiness and life satisfaction) since 1974 using 5-point verbal scales. Internationally, there is a tendency to use numerical scales rather than verbal scales, for example an 11-point scale ranging from 0 to 10 (Diener, Inglehart and Tay, 2012). SN has conducted three experiments in order to learn more about the consequences of changing from verbal to numerical response scales, about which numerical scales to use, and how to use them. First, in the Social Cohesion Survey in 2012 a split-half design was implemented, which included both a 5- and an 11-point scale. This experiment was conducted to see how changing from a verbal to a numerical scale would impact results. In the second experiment in the same year a three-way split design was used in the LISS-panel<sup>1</sup>, comparing different numerical 10- and 11-point scales. This experiment taught us more about which numerical scale to adopt. In the third experiment, conducted in the SN Web-panel, respondents rated their subjective well-being on a numerical scale and subsequently indicated which of the verbal categories corresponded best to their answer on the former question. Through this experiment we were able to define cut-off points and see whether demographically different groups interpret numerical scales in the same way.

In this report we discuss the results of these experiments. In the next section we provide a short overview of the literature on response scales. In section 3 we discuss the design and results from the first experiment in the Social Cohesion Survey. Section 4 describes the results of the LISS-panel experiment. The results of the SN Web-panel experiment are presented in section 5. We end with a discussion and some overall conclusions in section 6.

## 2 Literature

There is a wide range of literature on measuring subjective well-being (SWB). SWB refers to people's own cognitive and affective evaluations of their life (Diener, 2000) and is most often measured as happiness and/ or life satisfaction. However, due to a plethora of measures there is much debate about the optimal well-being measurement. A short literature overview on the main issues is provided in this section.

### 2.1 Numerical versus verbal response scales

In general, attitude measures are more reliable when they are more extensively labelled (Alwin and Krosnick, 1991). Therefore, fully labelled verbal scales should be more reliable than numerical scales. The trade-off concerns the number of response categories versus the number of verbal labels that can be included in the scale. That is, it is more difficult to label all answer categories when there are eleven rather than just five answer categories. However, verbal labels affect respondents'

---

<sup>1</sup> See [http://www.lissdata.nl/lissdata/About\\_the\\_Panel/General](http://www.lissdata.nl/lissdata/About_the_Panel/General) for more information about the LISS-panel.

scores especially when these labels do not divide the scale intervals approximately equally, leading to skewed scales (Wildt and Mazis, 1978). Otherwise, respondents can be expected to be able to interpolate the meaning of a numerical category solely based on the end labels.

In their meta-analysis Churchill and Peter (1984) do not find any difference between numerical versus verbal scales on scale reliability coefficients for a range of psychological studies. In addition, there is no relationship between labelling and scale reliability. That is, labelling all scale points versus only labelling end points of the scale does not increase reliability. In another meta-analysis based on 154 studies Peter and Churchill (1986) show a moderate correlation of 0.25 between measurement characteristics such as scale type and reliability.

Schwarz et al. (1991) show that the verbal labels not only affect the interpretation of numerical values, but also the values themselves can affect the interpretation of the scale. That is, scales ranging from “0” to “10” are interpreted differently than scales ranging from “-5” to “5”, even though the number of response categories are the same. Respondents rate their attitude differently when verbal labels are the same but the values differ. The authors recommend using positive numbers, since respondents are hesitant to assign a negative number to their attitude.

Numerical scales can be divided into anchoring scales, in which end labels are given, and self-anchoring scales, where no labels are given. Schifini D’Andrea and Maggino (2004) propose using anchoring scales since the presence of end labels unifies respondents’ interpretations of the scale.

As for verbal labels, a distinction can be made between unipolar and bipolar scales. Unipolar scales only measure one concept, whereas bipolar scales measure two opposite concepts. For example, a unipolar scale on happiness ranges from “not happy at all” to “very happy”, whereas the bipolar equivalent ranges from “very unhappy” to “very happy”. In a series of experiments Gannon and Ostrom (1996) show that unipolar scales are interpreted differently than bipolar scales concerned with the same subject. Moreover, bipolar scales make categories at the low end of the scale explicit and activate separate knowledge structures associated with the labels. In contrast, the low end of the scale is open to interpretation in unipolar scales. Therefore, in general bipolar scales are preferred.

## **2.2 Number of response categories**

In a study combining six surveys Andrews (1984) shows that compared to other survey characteristics, the number of scale categories has the largest influence on data quality. As more answer categories are used, validity tends to increase whereas residual errors decrease. Cox (1980) concludes in his literature review that more than nine response alternatives do not improve measurement anymore. Friedman and Amoo (1999) suggest that the number of scale points should depend on the subject. Thus, if people have more elaborate attitudes towards a subject there should be more answer categories.

Scherpenzeel and Saris (1993) recommend using an 11-point scale to measure satisfaction in particular. Cummins and Gullone (2000) recommend using 11-point scales rather than 5-point scales to measure subjective quality of life in order to increase scale sensitivity. That is, five answer categories provide too little variance. More answer categories do not decrease scale validity but increase sensitivity, because respondents are able to give more precise answers.

The OECD (2013) provides a set of guidelines to measure subjective well-being. Concerning response scales the OECD recommends using a “0-10 point numerical scale anchored by verbal labels which represent conceptual absolutes (such as completely satisfied/ completely dissatisfied). On balance, it is preferable to label scale interval-points (between the anchors) with numerical, rather than verbal, labels” (p. 14). This is an important guideline for Statistics Netherlands and would suggest using 11-point scales with bipolar verbal end labels.

### **3 Numerical versus verbal response scales: experiment in Social Cohesion Survey**

Statistics Netherlands has always used verbal 5-point scales to measure subjective well-being. In light of the international developments and recommendations by the OECD, we want to know whether and how changing to numerical 11-point scales affects results. For this reason, we conducted an experiment in the Social Cohesion Survey 2012, in which a split-half design was implemented assigning respondents randomly to either the old 5-point scale or the numerical 11-point scale. The results of this experiment can be used to test the comparability of both scales. This section describes the method and the results of this first experiment.

#### **3.1 Method**

##### *3.1.1 Data collection*

The Social Cohesion Survey 2012 conducted by Statistics Netherlands consists of questions on social contacts, participation, trust, and well-being. It included a split-half design in which two different response scales for questions on subjective well-being were tested. One version of the survey contains questions on happiness and life satisfaction using the original 5-point, verbally labelled scale, and the other version contains the same questions using a numerical 11-point scale ranging from “0” to “10”, where only the end points are labelled. Respondents were randomly assigned to either one of these two versions.

Data were collected using a sequential mixed mode design. People were sent an invitation and two reminder letters asking them fill out the questionnaire online (i.e. CAWI). Those who did not respond to this invitation were called and interviewed by phone (CATI) when a telephone number was available. When no telephone number was available people were interviewed face-to-face at their home (CAPI).

##### *3.1.2 Sample*

In total, 7 949 respondents of 15 years and older participated in the study (response rate of 61.6 per cent). Only respondents of 18 years and older are included in the analyses, resulting in a total number of respondents of 7 641. The version with the verbally labelled 5-point scales was distributed to 3 845 respondents, and the version with the numerical 11-point scales to 3 796 respondents. These two random samples are comparable in terms of sex, age, level of education, denomination, province, degree of urbanization, data collection mode, and data collection period. Therefore, we assume that any differences in the results are due to the different response scales.

### *3.1.3 Questions on subjective well-being*

The questions on subjective well-being used for the experiment are:

#### **Happiness-5**

To what extent do you consider yourself a happy person. Are you:

1. very happy,
2. happy,
3. neither happy nor unhappy,
4. not that happy,
5. or unhappy?

#### **Happiness-11**

On a scale from 0 to 10 can you indicate to what extent you consider yourself to be a happy person. A score of 0 refers to very unhappy and a 10 to very happy?

#### **Satisfaction-5**

To what extent are you satisfied with the life you currently lead. Are you:

1. extraordinarily satisfied,
2. very satisfied,
3. satisfied,
4. fairly satisfied,
5. or not that satisfied?

#### **Satisfaction-11**

On a scale from 0 to 10 can you indicate to what extent you are satisfied with the life you currently lead. A score of 0 refers to completely dissatisfied and a 10 to completely satisfied.

For the analyses, we recoded the scores for the happiness-5 and satisfaction-5 questions such that a higher score reflects a higher degree of happiness/ life satisfaction.

The labels of these 5-point scales are different for the questions on happiness and life satisfaction. For life satisfaction the labels are very asymmetrical whereas for happiness they are more symmetrical.

The verbal 5-point scale on happiness is bipolar (i.e., ranging from “unhappy” to “very happy”), whereas the life satisfaction scale is unipolar (i.e., ranging from “not that satisfied” to “extraordinarily satisfied”). The 11-point scales are bipolar scales with different, opposite end labels, both for the question on happiness and on life satisfaction.

## **3.2 Results**

First, we discuss the response distributions and the percentages of happy and satisfied people according to the various scales. Results are also specified for a number of relevant subgroups. Finally, the correlations are considered. Only weighted statistics of respondents aged 18 and older are included.

### *3.2.1 Distributions*

All four answer distributions violate the normality assumption according to the Kolmogorov-Smirnov test. When skewness and kurtosis are analysed, the high kurtosis of the 11-point scales stands out. Therefore, the 11-point scales seem more sensitive to normality violations. Very few respondents select a “0”, “1”, or “2” on the 11-point scales.

Respondents who use the 11-point scales are more likely to give the same rating to happiness and life satisfaction than participants who use the 5-point scales. Whereas 17.3 percent selects the same numerical category on the 5-point scales of happiness and life satisfaction, 31.3 percent does so on the 11-point scales. This is probably because the verbal labels clearly differ for happiness and life satisfaction.

The percentage of missing values is 0.9 percent on the happiness 5-point scale and 2.5 percent on the happiness 11-point scale. For life satisfaction, the percentages of missing values are 0.7 and 2.7 percent on the 5- and 11-point scales respectively. Both 5-point scales have significantly lower missing values than the corresponding 11-point scales ( $t_{\text{happiness}}(6195) = -5.48; p < 0.01$ ;  $t_{\text{satisfaction}}(5608) = -6.84; p < 0.01$ ).

**Table 1. Distributions of happiness and life satisfaction scales, 2012**

	Kolmogorov-Smirnov statistic	df	Skewness	Kurtosis
Happiness-5	0.36*	3787	-0.97	2.32
Happiness-11	0.24*	3695	-1.45	5.52
Satisfaction-5	0.24*	3796	-0.38	0.16
Satisfaction-11	0.22*	3682	-1.41	4.56

Note. \*  $p < 0.001$ .

**Table 2. Percentage of respondents per answer category, 2012**

5-point scales			
Happiness-5		Satisfaction-5	
	%		%
1. unhappy	0.6	1. not that satisfied	5.9
2. not that happy	3.4	2. fairly satisfied	11.1
3. neither happy, nor unhappy	12.7	3. satisfied	44.6
4. happy	65.8	4. very satisfied	31.2
5. very happy	17.5	5. exceptionally satisfied	7.2
11-point scales			
Happiness-11		Satisfaction-11	
	%		%
0. very unhappy	0.3	0. completely dissatisfied	0.3
1.	0.2	1.	0.3
2.	0.2	2.	0.3
3.	0.6	3.	0.9
4.	0.8	4.	1.0
5.	3.3	5.	4.1
6.	5.7	6.	8.0
7.	23.3	7.	24.8
8.	43.5	8.	39.4
9.	16.1	9.	15.1
10. very happy	6.2	10. completely satisfied	5.7

### 3.2.2 Percentages of happy and satisfied people

The results on the happiness and life satisfaction questions can be dichotomized in order to identify happy (unhappy) and satisfied (dissatisfied) people. Various cut-off points can be used. For the verbally labelled 5-point scales we used the dichotomies



currently applied at SN for the question on happiness, the two positive categories represent the happy people (consisting of “very happy” and “happy”), whereas for the life satisfaction question the three positive categories combine to represent satisfied people (consisting of “extraordinarily satisfied”, “very satisfied”, and “satisfied”). As for the numerical 11-point scales, the categories rated seven or higher are used, both for happiness and for life satisfaction. While conceptually, a score of “6” or higher would be considered “happy” or “satisfied with life” on an 11-point scale, it is possible that respondents consider a score of “7” or higher to represent “happy” or “satisfied with life”. Using a cut-off point of seven on the 11-point scales shows a higher consistency with the 5-point scales than the cut-off point of six. The reasons for choosing this cut-off point will be discussed in section 5. In table 3 the percentages of happy and satisfied people are presented for the different scales. The results show that the percentages of happy or satisfied people are significantly lower on the 5-point scales than on the 11-point scales ( $\chi_{\text{happiness}}^2(1) = 53.75; p < 0.01; \chi_{\text{satisfaction}}^2(1) = 6.71; p < 0.05$ ).

**Table 3. Percentages of happy and satisfied people across scale types, 2012**

Scale type	Proportion
	%
Happiness-5	83.3
Happiness-11, cut-off point 6	94.7
Happiness-11, cut-off point 7	89.0
Satisfaction-5	83.0
Satisfaction-11, cut-off point 6	93.1
Satisfaction-11, cut-off point 7	85.1

Next, we analysed the percentages of happy and satisfied people for various demographic subgroups based on differences in sex, age, ethnicity, and level of education. Differences are tested using logistic regressions. Overall, the results on the 5-point scales show lower percentages of happy and satisfied people than on the 11-point scales. In addition, there are more differences between groups on the 5-point than on the 11-point scales<sup>2</sup>.

#### *Sex*

There are no significant differences between men and women on any of the scales.

#### *Age*

There are no significant differences between age classes on the happiness and life satisfaction 11-point scales. The age group of 50-65 stands out on the 5-point scales: they score significantly lower than the two younger age groups of 18-35- and 35-50-year olds on both the happiness and the life satisfaction question, and on the 5-point satisfaction scale this age group scores lower than the over 65s.

#### *Ethnicity*

As for ethnicity, on the 11-point scales non-western immigrants score lower on happiness and life satisfaction than western immigrants and native Dutch people. On the 5-point scales non-western immigrants score lower on happiness and life

<sup>2</sup> Another study comparing the 5- and 11-point life satisfaction scales based on the Cultural Changes in the Netherlands Survey 2010 shows similar results and conclusions: the percentage of satisfied people is higher on the 11-point scale and the transformed mean is higher on the 11- than on the 5-point scale (van Beuningen, 2012).

satisfaction than native Dutch people. Western immigrants are also less satisfied than the native Dutch on the 5-point life satisfaction scale.

*Level of education*

Overall, people with the lowest level of education (solely primary education) are less happy or satisfied than people with a higher level of education. This is true for the 5-point scales. On the 11-point scales people with primary education only score lower than highly educated people with college and/ or university degrees. As for the 5-point happiness scale, people with pre-vocational education also score lower than people with a college or university degree. As for life satisfaction, people with a primary and junior general secondary education scored lower on the 11-point scale than people with a university education.

In conclusion, it is likely that group classifications in terms of well-being will change when SN opts for an 11-point scale.

**Table 4. Percentages of happy and satisfied people across scale types and demographic groups, 2012**

	Happiness-5	n	Happiness-11 cut-off point 6	Happiness-11 cut-off point 7	n
	%		%	%	
Total	83.3	3 670	94.7	89.0	3 605
<i>Sex</i>					
Male	82.2	1 809	94.8	89.4	1 768
Female	84.3	1 861	94.5	88.6	1 837
<i>Age</i>					
18 to 35 years	88.0	952	95.8	90.1	982
35 to 50 years	84.3	1 002	93.7	88.6	994
50 to 65 years	78.5	976	93.5	87.9	890
65 years or older	82.0	740	95.8	89.5	739
<i>Ethnicity</i>					
Dutch	84.2	2 960	96.2	90.5	2 849
Western	81.3	372	95.5	89.4	351
Non-western	77.0	339	83.0	77.8	405
<i>Level of education</i>					
Primary	71.5	331	90.1	81.9	312
Pre-vocational	80.8	671	94.1	88.6	686
Junior general secondary	84.8	1 571	95.4	89.2	1 529
Senior general secondary, senior vocational	86.8	681	95.7	91.9	701
Vocational college, university	88.7	361	97.5	92.0	319

	Satisfaction- 5	n	Satisfaction- 11 cut-off point 6	Satisfaction- 11 cut-off point 7	n
	%		%	%	
Total	83.0	3 679	93.1	85.1	3 595
<i>Sex</i>					
Male	82.6	1 813	92.7	85.8	1 765
Female	83.4	1 866	93.5	84.3	1 830
<i>Age</i>					
18 to 35 years	85.3	954	93.0	83.9	978
35 to 50 years	83.3	1 004	92.1	84.9	989
50 to 65 years	78.8	978	92.9	84.8	891
65 years or older	85.0	744	94.7	87.0	737
<i>Ethnicity</i>					
Dutch	84.3	2 966	94.7	87.0	2 846
Western	79.4	374	92.8	84.1	350
Non-western	75.0	339	81.4	72.0	399
<i>Level of education</i>					
Primary	72.6	332	87.9	76.2	312
Pre-vocational	80.5	672	92.2	84.6	687
Junior general secondary	85.3	1 573	92.9	84.4	1 524
Senior general secondary, senior vocational	84.9	682	95.6	89.9	698
Vocational college, university	86.1	364	97.2	89.1	317

Note. The original number of respondents is reported without weighting.

### 3.2.3 Correlations

As a final analytical step we analysed relationships of happiness and life satisfaction with related variables. Previous research shows that contacts with family members outside the household, with friends and acquaintances, and with neighbours are positively related to happiness and life satisfaction, as is one's self-reported health status (van Beuningen and Kloosterman, 2011; Mars and Schmeets, 2011). As expected, happiness and life satisfaction are significantly correlated with social contacts and self-reported health. The correlations are not significantly different for the various scales.

**Table 5. Correlations between subjective well-being (happiness and life satisfaction) and related constructs, 2012**

	Happiness-5	Happiness-11	Life satisfaction-5	Life satisfaction-11
Family contacts	0.14	0.13	0.12	0.14
Friends contacts	0.15	0.18	0.14	0.17
Neighbours contacts	0.14	0.14	0.12	0.16
Self-reported health status	0.39	0.40	0.42	0.39

Note. Reported correlations are significant ( $p < 0.05$ ).

#### **4 Number of response categories: experiment in LISS-panel**

In the previously described experiment a numerical 11-point scale was compared to a verbal 5-point scale. One of the assumed advantages of numerical 11-point scales is that they contain a neutral midpoint. It has been argued however, that this option may serve as an escape anchor for some respondents, and that 10-point scales are therefore preferable. Also, in the Dutch school system grades range from 1 to 10 with 5 denoting an insufficient result. So, for Dutch people the neutral midpoint on 11-point scales (i.e. the number 5 given that 11-point scales usually range from 0 to 10) may not signify a neutral position at all. To compare the use of 10- and 11-point scales an experiment was designed that included three different rating scales: (1) a 10-point scale, (2) an 11-point scale with the endpoints being defined by verbal labels, and (3) an 11-point scale with both the endpoints and the midpoint being defined by verbal labels. The last scale was added in order to test if and in what way labelling the midpoint of an 11-point rating scale would change people's ratings.

#### **4.1 Method**

##### *4.1.1 Data collection*

The experiment was conducted in October 2012 in the LISS-panel (Longitudinal Internet Studies for the Social Sciences). The LISS-panel consists of 5 000 households, comprising 8 000 individuals. The panel is based on a true probability sample of households drawn from the population register by SN. Because surveys in the panel are administered through the internet, households that could not otherwise participate are provided with a computer and internet connection.

Respondents were randomly assigned to one of three conditions that differed in the rating scales that were offered. In each condition participants were asked to subsequently rate their overall satisfaction with life, their satisfaction with a number of life domains (e.g. financial situation of the households, personal relationships) and their overall level of happiness. At the end of the questionnaire they were asked to rate the previous questions on a number of aspects, such as difficulty, clarity and enjoyability.

##### *4.1.2 Sample*

In total 6 874 panel members of 16 years and older were offered participation in the questionnaire, 5 687 of whom started answering the questionnaire and 5 685 completed the entire questionnaire, bringing the overall response rate to 83 percent. Only respondents of 18 years and older are included in the analyses, resulting in a total number of respondents of 5 556. The version with the 10-point scale was distributed to 1 821 respondents, the version with the 11-point scale endpoints labelled to 1 891 respondents and the version with the 11-point scale with both the endpoints and midpoint labelled to 1 844 respondents. Given the random procedure

and the large sample, it is safe to assume that any differences in the results are due to the different response scales.

#### *4.1.3 Questions on subjective well-being*

The questions on subjective well-being reported on here are:

**Happiness 1-10:**

On a scale from 1 to 10 can you indicate to what extent you consider yourself to be a happy person? A score of 1 refers to completely unhappy and a 10 to completely happy.

**Happiness 0-10:**

On a scale from 0 to 10 can you indicate to what extent you consider yourself to be a happy person? A score of 0 refers to completely unhappy and a 10 to completely happy.

**Happiness 0-5-10:**

To what extent do you consider yourself a happy person? 0 means completely unhappy and 10 completely happy. A score of 5 means neither unhappy nor happy.

**Life satisfaction 1-10:**

On a scale from 1 to 10 can you indicate how satisfied you are with the life you lead at the moment? A score of 1 refers to completely dissatisfied and a 10 to completely satisfied.

**Life satisfaction 0-10:**

On a scale from 0 to 10 can you indicate how satisfied you are with the life you lead at the moment? A score of 0 refers to completely dissatisfied and a 10 to completely satisfied.

**Life satisfaction 0-5-10:**

On a scale from 0 to 10 can you indicate how satisfied you are with the life you lead at the moment? A score of 1 refers to completely dissatisfied and a 10 to completely satisfied. A score of 5 means neither dissatisfied nor satisfied.

## **4.2 Results**

First, we discuss the response distributions and means for the various scales, both for the overall sample and for a number of relevant subgroups. Next, we show the reaction times to the questions in the different conditions. Finally, respondents' evaluations of the various scales are discussed.

### *4.2.1 Distribution and means*

Tables 6 and 7 show the percentage frequency distribution for life satisfaction and happiness in each condition. Mean scores on life satisfaction and happiness did not differ between conditions. Also, no differences were found between conditions on the percentage of people scoring 7 or above (the reasons for choosing this cut-off point will be discussed in section 5). Use of the different rating scales was not dependent on age, sex, education level or ethnicity with one exception: the interaction between condition and ethnicity was significant for the mean score on overall life satisfaction. Both first and second generation immigrants had lower mean scores than the native Dutch on both the 10-point and 11-point rating scale without midpoint. On the 11-point rating scale with midpoint label first generation immigrants had lower mean scores than the native Dutch and second generation immigrants. However, the findings with regard to ethnicity should be regarded with

caution given the low number of first and second generation immigrants in the sample (approximately 110 first and 110 second generation immigrants and 1600 native Dutch people in each condition).

**Table 6. Percentage frequency distribution and means for overall life satisfaction in each condition**

	Response scale			sig.
	1-10	0-10	0-5-10	
	n = 1821	n = 1891	n = 1844	
<i>Score</i>				
0	-	0	0	
1	0	0	0	
2	0	1	1	
3	1	1	1	
4	2	2	2	
5	4	4	6	
6	7	7	6	
7	22	24	23	
8	41	41	38	
9	18	15	18	
10	4	5	5	
Mean (sd)	7,6 (1,4)	7,6 (1,4)	7,5 (1,5)	ns
% $\geq 7$	85	85	83	ns

**Table 7. Percentage frequency distribution and means for happiness in each condition**

	Response scale			sig.
	1-10	0-10	0-5-10	
	n = 1864	n = 1928	n = 1893	
<i>Score</i>				
0	-	0	0	
1	0	0	0	
2	0	1	0	
3	0	1	1	
4	1	1	1	
5	3	4	5	
6	7	7	7	
7	23	21	20	
8	37	39	38	
9	22	21	22	
10	5	5	5	
Mean (sd)	7,7 (1,3)	7,7 (1,4)	7,7 (1,4)	ns
% $\geq 7$	87	87	85	ns

#### 4.2.2 Reaction times

The reaction time to a question can be regarded as a measure of the difficulty of that question. A quick response may indicate that the question was easy to answer whereas a slow response may indicate that the question was difficult to answer. Because each question was presented on a new screen it was possible to calculate reaction times to each question. First, reaction time outliers were identified for each question using the interquartile range (IQR):

Lower threshold:  $Q1 - 1.5 * IQR$

Higher threshold:  $Q3 + 1.5 * IQR$

Based on this definition approximately 5 percent of all reaction times were considered outliers. After outlier removal, reaction times were compared for the three conditions on the two well-being questions (see Table 8).

**Table 8. Reaction times (in seconds) for overall life satisfaction and happiness in each condition**

	Response scale			sig.
	1-10	0-10	0-5-10	
Life satisfaction	8.53	8.13	9.04	$p < .001$
Happiness	6.77	6.54	6.82	$p < .05$

The reaction times to both the life satisfaction and the happiness question were dependent on the condition. For life satisfaction, response was faster when an 11-point rating scale without midpoint label was used than when a 10-point scale was used. Response was slowest when an 11-point scale with midpoint label was used. For happiness, again, response was faster when an 11-point rating scale without midpoint was used than when an 11-point scale with midpoint was used. However, no differences in reaction time were found between the 10-point scale and the 11-point scale without midpoint label nor between the 10-point scale and the 11-point scale with midpoint label.

The reaction time results with regard to life satisfaction suggest that it was easier for participants to use the 11-point rating scale without the midpoint labelling than the 10-point rating scale which in turn was easier to use than the 11-point rating with midpoint labelling. However, it should be kept in mind that the reading instructions were longer for the latter scale which could easily explain the difference between 10-point and 11-point scale with midpoint labelling. Furthermore, one of the so-called core questionnaires that respondents answer every year contains questions on well-being that use 11-point rating scales with endpoint labelling. This means that respondents are used to this type of scale and this in turn may explain the fact that it was easier for them to use the 11-point rating scale without midpoint labelling than the 10-point rating scale. It is also important to note that differences in reaction times between conditions disappeared after two questions suggesting that respondents quickly got used to the scales they were asked to use. In the last question, about happiness, the labels used for the endpoints changed which probably caused the differences in reaction time between the three conditions. Again, the shorter reaction time to the question for respondents who used the 11-point rating scale without the midpoint labelling is probably due to the fact that this question is also part of the core questionnaire and respondents are therefore familiar with it.

#### 4.2.3 Respondents' evaluation of the scales

At the end of the questionnaire respondents were asked to evaluate the well-being questions on five aspects, three of which were relevant to our experiment: difficulty,

clarity and enjoyability. Evaluations were made on a numerical 5-point scale of which the end points were labelled “certainly not” and “certainly yes”. Differences between conditions were found with regard to clarity and enjoyability (see Table 9). Well-being questions were evaluated as clearer and more enjoyable when using 10 point rating scales than when using 11-point rating scales with a midpoint label. No differences were found between 10-point rating scales and 11-point rating scales without a midpoint label.

**Table 9. Percentage of respondents who rated the well-being questions as not difficult, clear and enjoyable in each condition**

	Response scale			sig.
	1 - 10	0-10	0-5-10	
Difficult: certainly not	66	66	64	ns
Clear: certainly yes	63	61	59	p < .05
Enjoyable: certainly yes	32	32	28	p < .05

### 5 Respondents’ scale interpretations: experiment in SN Web panel

While conceptually, a score of “6” or higher on a 10- or 11-point numerical scale would be considered “happy” or “satisfied with life”, it is possible that respondents consider a higher score to represent these concepts. To shed light on this issue a third experiment was designed that asked respondents to rate their overall level of happiness and life satisfaction on a 10-point numerical scale and then asked them to indicate which label on a 5-point verbal response scale was its equivalent.

#### 5.1 Method

##### 5.1.1 Data collection

The third experiment was conducted in June 2013 in the SN Web panel. The SN Web panel was formed at the end of December 2012 as part of the pilot project ‘Web panel’. The aim of this project was to collect information about the challenges and possibilities of using a web panel to collect statistical information in the near future. People who had participated in the National Mobility Survey, who had indicated that they would be willing to participate in other research and who had access to the internet, were approached for participation in the SN Web panel. At the beginning of 2013 this panel consisted of approximately 1 200 members. At the beginning of every month, starting in January 2013, panel members were asked to fill out a questionnaire online. The June 2013 questionnaire asked about the following subjects: preventive health care, organ donation, well-being, voluntary work, political participation and trust.

##### 5.1.2 Sample

In total 1 162 panel members aged 16 years and older were offered participation in the questionnaire, 989 of whom started answering the questionnaire and 976 completed the entire questionnaire, bringing the overall response rate to 84 percent. Only respondents of 18 years and older are included in the analyses, resulting in a total number of respondents of 970.

##### 5.1.3 Questions on well-being

The following questions on subjective well-being were used in this experiment:



### ***Happiness-10***

On a scale from 1 to 10 can you indicate to what extent you consider yourself to be a happy person? A score of 1 refers to completely unhappy and a 10 to completely happy.

### ***Happiness-5***

You just rated your level of happiness with a [RESPONSE TO FORMER QUESTION]. Can you indicate to which of the following categories this corresponds best:

1. very happy,
2. happy,
3. neither happy nor unhappy,
4. not that happy,
5. or unhappy?

### ***Life satisfaction-10***

On a scale from 1 to 10 can you indicate how satisfied you are with the life you lead at the moment? A score of 1 refers to completely dissatisfied and a 10 to completely satisfied.

### ***Life satisfaction-5***

You just rated your level of life satisfaction with a [RESPONSE TO FORMER QUESTION]. Can you indicate to which of the following categories this corresponds best:

1. extraordinarily satisfied,
2. very satisfied,
3. satisfied,
4. fairly satisfied,
5. or not that satisfied?

## **5.2 Results**

In a first analytical step we compare the results of this experiment to the results from the Social Cohesion Survey (section 5.2.1). Next, in section 5.2.2 results are given on the respondents' interpretation of the scales, both for the total sample as for a number of demographic subgroups.

### ***5.2.1 Comparison of results***

Tables 10 and 11 show the results of the comparisons between the Web panel and the Social Cohesion results. A number of things were different in the way the questions were asked: first of all, in the Social Cohesion Survey an 11-point scale was used, and in de Web panel a 10-point scale was used. Second, respondents in the Social Cohesion Survey were offered the possibility to refuse to answer the question, or to say that they don't know the answer. In total, almost 100 respondents didn't answer the question. Respondents in the Web panel did not have these possibilities and were obliged to answer the questions.

For happiness the response distribution of was very similar in both surveys. For life satisfaction questions there were slightly more differences, but still the results are quite similar.

**Table 10. Comparison of results on happiness**

	Social Cohesion		Web panel	
	%	n	%	n
<i>Score</i>				
0	0	10	-	-
1	0	5	0	2
2	0	7	0	2
3	1	20	1	7
4	1	29	1	8
5	3	118	2	19
6	5	200	5	47
7	23	868	25	241
8	44	1662	46	444
9	16	622	18	175
10	6	233	3	25

**Table 11. Comparison of results on life satisfaction**

	Social Cohesion		Web panel	
	%	n	%	n
<i>Score</i>				
0	0	12	-	-
1	0	18	0	1
2	0	12	0	4
3	1	32	0	3
4	1	36	1	9
5	4	143	2	22
6	7	282	7	69
7	25	933	27	259
8	40	1493	43	421
9	16	588	17	166
10	6	223	2	16

### 5.2.2 Respondents' interpretation of the scales

Tables 12 and 13 show the relationship between the numeric and verbal scale as interpreted by the total sample of respondents. For happiness for example we see that out of all respondents who rate their level of happiness with a '10', 3 would say that they are happy and 22 would say that they are very happy. There are a few strange responses, one respondent for example rates his or her level of happiness with a '1', and claims to be very happy. In total, 4 of such responses are given, and they are treated as outliers<sup>3</sup>.

<sup>3</sup> These responses are marked with a \* in tables 12 and 13.

**Table 12. Respondents' interpretation of the happiness 10-point scale (absolute numbers)**

	Unhappy	Not that happy	Not happy, not unhappy	Happy	Very happy
1	1	0	0	0	1*
2	0	1	0	1*	0
3	2	5	0	0	0
4	1	6	1	0	0
5	0	11	8	0	0
6	0	9	30	8	0
7	0	0	62	179	0
8	0	0	8	402	34
9	0	0	0	66	109
10	0	0	0	3	22

Note. Responses marked with a \* are treated as outliers.

**Table 13. Respondents' interpretation of the life satisfaction 10-point scale (absolute numbers)**

	Not that satisfied	Fairly satisfied	Satisfied	Very satisfied	Extraordinarily satisfied
1	0	0	0	0	1*
2	3	0	0	1*	0
3	3	0	0	0	0
4	8	1	0	0	0
5	9	11	2	0	0
6	11	33	25	0	0
7	0	36	208	15	0
8	0	2	177	240	2
9	0	0	4	134	28
10	0	0	2	4	10

Note. Responses marked with a \* are treated as outliers.

In order to determine which number on the numerical scale represents being happy, we look at the verbal category that is chosen most often by respondents for each numeric score. So a score of '10' on happiness corresponds to being very happy, since this combination of scores is indicated most by respondents. This way, we find that for happiness, a score of 1 corresponds to being unhappy, scores 2 to 5 correspond to being not that happy, a score of 6 corresponds to being neither happy nor unhappy, scores 7 and 8 correspond to being happy and finally scores 9 and 10 correspond to being very happy. For life satisfaction, the results differ slightly: there is no valid response for a score of 1, scores 2 to 4 correspond to being not that satisfied, scores 5 and 6 to being fairly satisfied, a score of 7 corresponds to being satisfied, scores 8 and 9 to being very satisfied and a score of 10 finally to being extraordinary satisfied.

From these results we conclude that people do not seem to interpret a score of 6 as a clear positive option, although it is on the positive side of the scale. This leads to the conclusion that we cannot use a score of 6 as a cut-off point for determining whether someone is happy or satisfied with life. Since we want to use the same classification for both happiness and life satisfaction, we combined the results for both topics into the classification presented in table 14.

**Table 14. Classification of the 1-10 scale for happiness and life satisfaction**

	Happiness	Life satisfaction
1 to 4	Not happy	Not satisfied
5, 6	Neither happy nor unhappy	Neither satisfied nor dissatisfied
7, 8	Happy	Satisfied
9, 10	Very happy	Very satisfied

When looking at results for a number of subgroups, we find that there are no significant differences in how people belonging to different groups interpret the well-being scales. The subgroups that we explored are: sex, age, level of education, income, and health status. Unfortunately there was insufficient data to look at respondents' ethnicity (native Dutch and western/non-western immigrants).

## 6 Reflection and conclusions

The first experiment demonstrated that no meaningful differences were found in correlations between the 5- and 11-point scales and constructs related to happiness and life satisfaction. This means that both scales measure the same construct. However, the experiment also showed several differences between the two scales. First, applying cut-off points we find lower percentages of happy and satisfied people when using 5- than when using 11-point scales. This means that introducing 11-point scales will result in a deviation from the trend as new statistics will not be comparable to previous years. Second, the 5-point scales have fewer missing values than the 11-point scales. This could imply that respondents find it more difficult to rate their happiness and life satisfaction on an 11-point numerical scale with verbal end labels than on a 5-point verbal scale.

The second experiment compared three different numerical scales (10- and 11-point scales with endpoint labelling and an 11-point scale with both midpoint and endpoint labelling) and found very few differences between these scales. Mean scores on life satisfaction and happiness did not differ between scales. Also, no differences were found between scales on the percentage of people scoring 7 or above, nor were there relevant differences in how people belonging to demographically different groups used the scales. Some reaction time differences were found between the scales, but these were probably due to differences in instruction length and familiarity with the scales. However, well-being questions were evaluated as more clear and more enjoyable when using a 10 point-rating scale than when using an 11-point rating scale with a midpoint label. Based on this finding and the fact that Dutch people are familiar with 10-point rating scales SN decided to use 10 point-rating scales to measure subjective well-being in the future. Internationally, 11-point scales are often used. But 'deviation' from the international standard should not pose any problems given the fact that this experiment showed that 10- and 11-point scales are highly comparable.

The third experiment allowed us to determine cut-off points and to see whether demographically different groups interpret numerical scales in the same way. This experiment showed that people do not interpret a score of 6 as a clear positive option, although it is on the positive side of the scale. Instead, a cut-off point of 7 is indicated for both happiness and satisfaction with life. No significant differences were found in how people belonging to different groups interpret the well-being scales.

## References

- Andrews, F. M., 1984, Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach. *The Public Opinion Quarterly*, 48 (2), 409-442.
- Beuning, J., van, 2012, Analyse responschalen van tevredenheid met het leven en inkomen uit Culturele Veranderingen 2010. Internal report. Statistics Netherlands, Den Haag/Heerlen. BPA: SAL-2012-H5. (in Dutch)
- Beuning, J., van, and R. Kloosterman, 2011, Subjectief welzijn: welke factoren spelen een rol? *Bevolkingstrends*, 59 (4), 85-92. (in Dutch)
- Churchill, G. A., Jr. and J. P. Peter, 1984, Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis. *Journal of Marketing Research*, 21 (4), 360-375.
- Cummins, R.A. and E. Gullone, 2000, Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. *Proceedings, Second International Conference on Quality of Life in Cities*, 74-93. Singapore: National University of Singapore.
- Diener, E., 2000, Subjective Well-Being. *The Science of Happiness and a Proposal for a National Index*. *American Psychologist*, 55 (1), 34-43.
- Diener, E., R. Inglehart, L. Tay ,2012, Theory and Validity of Life Satisfaction Scales. *Social Indicators Research*, in press.
- Gannon, K. M. and T. M. Ostrom, 1996, How Meaning is Given to Rating Scales: The Effects of Response Language on Category Activation. *Journal of Experimental Social Psychology*, 32, 337-360.
- Mars, G. and H. Schmeets, 2011, Meer sociale samenhang, meer geluk? *Bevolkingstrends*, 59 (3), 39-44 (in Dutch).
- OECD, 2013, *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing. Available through <http://dx.doi.org/10.1787/9789264191655-en>
- Peter, J. P. and G. A. Churchill Jr., 1986, Relationships among Research Design Choices and Psychometric Properties of Rating Scales: A Meta-Analysis. *Journal of Marketing Research*, 23 (1), 1-10.
- Scherpenzeel, A. C. and W. E. Saris, 1993, The Evaluation of Measurement Instruments by Meta-Analysis of Multitrait-Multimethod Studies. *Bulletin de Methodologie Sociologique* 39, 3-19.
- Schifini D'Andrea, S. and F. Maggino, 2004, Different scales for different survey methods. *Società Italiana di Statistica*, 339-350.
- Schwarz, N., B. Knäuper, H.-J. Hippler, E. Noelle-Neumann, and L. Clark, 1991, Rating Scales: Numeric Values May Change the Meaning of Scale Labels. *The Public Opinion Quarterly*, 55 (4), 570-582.

Stiglitz, J. E., A. Sen en J. P. Fitoussi, 2009, Report by the Commission on the Measurement of Economic Performance and Social Progress, CMEPSP. Available through [www.stiglitz-sen-fitoussi.fr](http://www.stiglitz-sen-fitoussi.fr)

Wildt, A. R. and M. B. Mazis, 1978, Determinants of Scale Response: Label Versus Position. *Journal of Marketing Research*, 15 (May), 261-267.