HAPPINESS - General Target Variable Report (GVR)

1. General Information

Target variables T_HAPPY_11 and T_HAPPY_DISTRIB describe the respondent's self-declared level of happiness. The source values capturing the respondent's level of happiness are expressed in different scales. We propose to transform original (source) scales into the two common (target) variables with: 11-point scale (T_HAPPY_11) and cumulative distribution scale in percentages from 0 to 100 (T_HAPPY_DISTRIB).

We propose to use the harmonized target variable on happiness with control variables that capture special features of the source variables with regard to response scales: (a) scale length (C_HAPPY_LENGTH), (b) direction of the scale (C_HAPPY_ASCEND) and (c) polarity (C_HAPPY_POLARITY).

The target variable report for T_HAPPY is accompanied by the following Excel documents:

- The Detailed Variable Report (DVR) T_HAPPY_DVR_SDR2.xlsx. DVR Excel files in SDR2 systemize all information about source variables that were used for harmonization into a given target variable of the SDR2 database;
- The Crosswalk Table (CWT): T_HAPPY_CWT_SDR2.xlsx. CWT Excel files in SDR2 contain details about mapping of source values to target values.

Table 1.1. HAPPINESS: Description of the target, source, and control variables

	Variable description	**Variable name**	**Variable values***
Target variables	Level of happiness (11-point scale)	T_HAPPY_11	0 = Lowest degree 10 = Highest degree
	Level of happiness (distribution-preserving scale)	T_HAPPY_DISTRIB	0 = Lowest percentile point in distribution 100 = Highest percentile point in distribution
Source variables			See: T_HAPPY_DVR_SDR2.xlsx and T_HAPPY_CWT_SDR2.xlsx
Control variables	Source scale length	C_HAPPY_LENGTH	3= 3-point scale 4 = 4-point scale 5 = 5-point scale 7 = 7-point scale 10 = 10-point scale 11 = 11-point scale
	Source values: scale direction	C_HAPPY_ASCEND	0 = Descending 1 = Ascending
	Source values: scale polarity	C_HAPPY_POLARITY	1 = unipolar (happy) 2 = bipolar (happy/unhappy)

[a] Missing values are assigned according to the SDR2 missing codes schema, provided in the Appendix.

2. Survey Projects

Source variables that we used for T_HAPPY appear in 8 international survey projects: CB, EB, EQLS, ESS, EVS, ISSP, LB, WVS, 45 waves and 1035 national surveys. The data cover 115 countries and years from 1975 to 2017.

3. General Rules and Procedures

3.1 Source data description

The majority of the source variables containing information on HAPPINESS question general state of happiness of the respondents[1].

Most international survey projects ask about happiness in general with some variations: e.g.:

CB/2010-2015: Overall, how happy would you say you are? Please use this CARD, where code '1' corresponds to "Extremely unhappy" and code '10' corresponds to "Extremely happy;" please choose any number between 1 and 10, which best reflects your opinion.
ESS/1-8: Taking all things together, how happy would you say you are?
EQLS/1-3: Taking all things together on a scale of 1 to 10, how happy would you say you are?
WVS/1-6: Taking all things together, would you say you are: 1|Very happy 2|Quite happy 3|Not very happy 4|Not at all happy

The question wordings used to create source variables can be divided into two groups: a) asking respondents "how happy (would you say you are)"; b) asking respondent to declare whether he/she is "very happy, fairly happy, not too happy" or "very happy, fairly happy, not very happy, not at all happy".

There were different types of scales used in source variables, varying in length between 3 and 11 points.

Table 3.1 HAPPINESS: Types of scales

Length of scale	Direction of scale	
	Descending	**Ascending**
3	EB/3, 5, 8, 9, 11, 18, 19, 24, 25	
4	EVS/1-4; ISSP/1991, 1998, 2001, 2002, 2007, 2008; LB/2002, 2008; WVS/1-6	
5		ISSP/2011
7	ISSP/2002, 2011, 2012	
10		CB/2010-2015, EQLS/1-3
11		ESS/1-8

[1] For more details, see the Detailed Variable Report.

3.2 Rules of transformation of the source variable into target variables

To construct the 11-point scale and distribution-preserving target variables, we first create preparatory scales. This involves recording the values of the source scales using the consecutive numbers *k*, where *k* ranges from 1 to n. The value 1 of the preparatory scale corresponds to the lowest happiness, and higher scores correspond to higher happiness (ascending direction). Each preparatory scale is of the same length as the source scale it was derived from.

3.2.1 Constructing the 11-point target scale

To construct the 11-point target scale, we use the preparatory scales and assign scores to them in the interval from 0 to 10, according to the following linear transformation:

$$l(k) = \frac{10}{n*2} + (k - 1) * \frac{10}{n}$$

where: $l(k)$ is a target score corresponding to the preparatory score *k*, and *n* is the number of *k*-values.

This process involves "stretching" preparatory (and thus, source) scales that have fewer than 10 points and keeping 11-point scales as such. Figure 3.2.1 and Table 3.2.1 depict transformations using this type of rescaling.

Figure 3.2.1. Transformation of source values into the target 11-point scale with 0 to 10 range

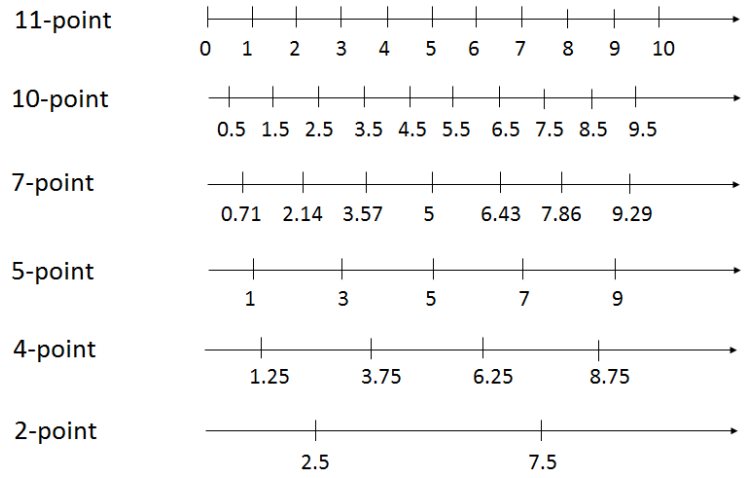

Table 3.2.1. Creating the 11-point scale (from 0 to 10), with median and mean values 5, and minimized inter-scale differences in the variability

Source scale length	Recodes	Median Mean	Average of absolute deviations	Variance	Standard deviation
11-point	0,1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0	5.0	2.72	10.00	3.16

10-point	0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5	5.0	2.50	8.25	2.87
7-point	0.7, 2.1, 3,6, 5.0. 6,4, 7.9, 9,3	5.0	2.46	8.25	2.87
5-point	1.0, 3.0, 5.0, 7.0, 9.0	5.0	2.40	8.00	2.83
4-point	1.25, 3.75, 6.25, 8.75	5.0	2.50	7.81	2.80
3-point	1.7, 5.0, 8.3	5.0	2.20	7.26	2.69
2-point	2.5, 7.5	5.0	2.50	6.25	2.50

3.2.2 Constructing the distribution-preserving target scale

To construct the distribution-preserving target scale, we take into account respondents' position in the distribution of reported life satisfaction values in a given national sample. For an *n*-point preparatory scale, for values *k* that range from 1 to *n,* where X_i is the percent distribution of the variable in sample *s*, *k* is recoded to:

$$k = \sum_{i=1}^{k-1} X_i + \frac{X_k}{2}$$

The distributional score for the answer option *k* is the sum of percentiles of all previous answer options up to *k-1* plus half of the percentile of the answer option *k*.

For a given sample, each scale point of the distribution target scale corresponds to the midpoint of the cumulative distribution of scores *k* (see Table 3.2.2). Put differently, the scores of the distributional target scale are percentiles that indicate what share of respondents within a national sample reports the same or lower value than the individual. The target variable is computed using unweighted samples.

Table 3.2 illustrates how we transform **preparatory** variables (which recode **source** variables' values in ascending direction) with 5 response options into the distribution-based target variable.

Table 3.2. Example of the distribution-based transformation of 5-point preparatory variables into T_HAPPY_DISTRIB.

Preparatory variable values, based on source values k	**Percentage distribution** X_k	**Cumulative percentage distribution** $\sum_{i=1}^{k} X_i$	**Interval** $\sum_{i=1}^{k-1} X_i$	**Interval lower bound plus interval midpoint** $\sum_{i=1}^{k-1} X_i + \frac{X_k}{2}$	**Target value (rounded to integer)**
1 = lowest happiness	10.68	10.68	0	= 10.68/2 = 5.34	5
2	32.75	43.44	10.68	= (10.68 + 32.75)/2 = 27.05	27

3	32.11	75.55	43.44	= (43.44 + 32.11)/2 = 59.49	59
4	21.69	97.23	75.55	= (75.55 + 21.69)/2 = 86.39	86
5 = highest happiness	2.77	100	97.23	= (97.23 + 2.77)/2 = 98.61	99

Missing values and different situations that warrant to be treated as missing data are coded according to the SDR2 missing codes schema, provided in Table A.1 in the Appendix.

3.3. Methodological variables that accompany T_HAPPY

The main variation in source questions about trust in parliament stems from differences in response scales. We provide three harmonization control variables that store specific features of the source variables (see Table 1.1):

1.

- *Scale length* (C_HAPPY_ LENGTH): Number of points/options in the source response scale. Values: 3, 4, 5, 7, 10, 11.
- *Scale direction* (C_HAPPY_ASCEND): Coded 1 if the source scale is ascending, i.e., responses are ordered from least happy to very happy.
- Scale polarity (C_ HAPPY_POLARITY): Coded 1 if the source scale is unipolar, i.e., the concept is measured along one dimension (from not at all satisfied with life to very satisfied) and 2 in the case of item-specific bipolarity (e.g., from unhappy to happy).

4. Special cases

- Some variables containing information about respondents self-declared happiness **were excluded** from our final list of target variables as the question wording did not fit the target concept, e.g. asking about frequency rather than intensity of feeling happy.

  For example:

  In EB/58.2 and EB/64.4 the questions are about how things have been: "*In the past month, have you felt happy?"* and "*These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks?*"

  EB/56.1 asks: "*Would you say that you have not at all, no more than usual, rather more than usual, much more than usual...? Been feeling unhappy and depressed*".

- ISSP/2011 provides a separate variable for China (`cn_v5`), as Chinese respondents were asked a slightly different question, i.e. "Overall, do you think that your life is happy? Very unhappy, Quite unhappy, I'm not sure, Happiness is not happiness, Quite happy, Very happy". This question differs from those asked in other ISSP 2011 countries in question wording, response scale and order of answer options. The question common across the other ISSP/2011 countries is: `v5` "If you were to consider your life in general these days, how happy or unhappy would you say you are, on the whole? Completely happy, Very happy, Fairly happy, Neither happy nor unhappy, Fairly unhappy, Very unhappy, Completely unhappy". We include both ISSP/2011 variables for the SDR2 target variable.

Appendix: Codes for missing values in SDR2

In the SDR database v.2 we identify different situations that warrant to be treated as missing data. Table A.1 lists all SDR2 missing value codes:

Table A.1. Codes for missing values in SDR2

SDR tag [a]	**SPSS (STATA) codes**	**Label**
		Standardized source codes for missing values
DK	-1 (.a)	Don't know
NA	-2 (.b)	No answer
REF	-3 (.c)	Refusal
DU	-4 (.d)	Don't understand the question
DNR	-5 (.e)	Any combination of DK, NA, REF, DU
INAP	-6 (.f)	Inapplicable
NEC	-7 (.g)	Not elsewhere classified
		SDR created codes for missing values
UNFIT	-8 (.h)	Source value does not fit to target
ERR	-9 (.i)	Errors in source data and undocumented source values
COMBI	-10 (.j)	Different missing codes on multiple sources taken for a target
CINAP	-11 (.k)	For control variables only: inapplicable
INSUF	-12 (.l)	For survey: Insufficiently defined response categories
QNA	-20 (.t)	For survey: Question not available

[a] Abbreviations for the labels corresponding to the SDR2 codes for missing values. These tags are used in the Crosswalk Table (CWT) files (Excel) that accompany documentation of SDR2 target variables.

In exceptional situations when codes for missing data listed in Table A.1 cannot be used, we apply a system missing <null> value.