

## 3 METHODS FOR MEASUREMENT OF HAPPINESS

Wim Kalmijn<sup>1</sup>

---

- 3/1 [Measurement of happiness in Social Science Research](#)
  - 3/2 [Standard method of measuring happiness in practice](#)
  - 3/3 [Happiness measured as a discrete variable at the ordinal level of measurement.](#)
  - 3/4 [Happiness research at the individual vs. the collective level](#)
  - 3/5 [Cumulative frequency distribution of happiness in a sample](#)
  - 3/6 [Index numbers for the statistical distribution of happiness](#)
  - 3/7 [Non-standard methods](#)
- 

### 3/1 Measurement of happiness in Social Science Research

Happiness in this context is defined as 'the degree to which an individual judges the overall quality of his life-as-a-whole positively', or in short: how well one likes the life one lives. In this way, happiness belongs to a wider class of subjective appraisals of life, which is usually referred to as 'subjective well-being' (SWB) or 'life satisfaction'. (cf. Chapter 2). We will confine this description to the measurement of happiness, but the following text mutatis mutandis equally applies to other forms of subjective well-being that are measured in a similar way.

Given the above definition of happiness, the obvious way to measure it is to ask the individual to give his or her opinion on one's own happiness situation.

The measurement of the happiness of a particular person as such is the objective of measurement in exceptional cases only; it occurs in psychological practice incidentally, but such situations have their own specific requirements that are not dealt with here.

Sociologists, however, are always interested in the happiness of (members of) collectivities, e.g., of the citizens of a nation, but yet they cannot get away from starting their study with measuring happiness at the individual level (Section 3/4). Moreover, they are usually hardly interested in happiness as such, but only in the association of happiness with one or more other variables. These other variables are

---

<sup>1</sup> Text May 2015

usually referred to as “correlates” or as “conditions”; the latter term is in particular in use if the observed association with happiness is supposed to be a causal one. In the case the investigation concerns “trends”, the other variable is time.

## 3/2 Standard method of measuring happiness in practice

### *Survey studies*

Measuring happiness for social scientific purposes is usually realized as a part of a ‘survey’ in which a lot of people answer the same questions, either in face-to-face interviews or on questionnaires presented on the web. The way questions are presented to the respondents is adjusted to this technique, and that means typically that questions are answered selecting from a limited number of response options (so called ‘closed’ questions). We will refer to this method as the *standard method*, which implies that also *non-standard methods* exist; these will be discussed in Section 3/7 of this chapter.

### *Survey questions*

To all members of a sample that is considered to be representative for the target population of the study, one or more closed questions are presented with a limited number of response options. This combination of the question and all response options together is referred to as a *primary scale of happiness measurement*. In the *World Database of Happiness* (further abbreviated as WDH), some one thousand such scales have been gathered in the so-called “Measures of Happiness” collection. In the Happiness Research literature they are often referred to as “*items*”.

### *Examples of survey questions on happiness*

An example of such a primary scale is the combination of the *single question*:

“Taking all things together, would you say you are ...? “

with four *response options*:

- very happy
- quite happy
- not very happy
- not at all happy.

These options are ordered either in ascending or – as in the above example – in descending order of (subjective) happiness intensity; this order should be unambiguous. The respondent is asked to tick the one out of the four response options he feels to be the most appropriate, or sometimes the least inappropriate.

The above measure of happiness is a standard in the studies of the “World Values Survey”; in the WDH it is coded as [O-HL-u-sq-v-4-a](#). The meaning of the symbols in such codes is described in Section 4 of this chapter. This measure, as most others, is, or at least should be extended in the survey with the options “Don’t know” and/or “No Answer” (notation “DK/NA”), but these are ignored in this context.

In Table 3.1 we mention this and two other items as examples; they are used frequently in Dutch surveys for time series over the period 1975 -2012.

The response options, which are also referred to as “*ratings*”, may be presented as text (“*verbal scales*”), as in the example, or as digits (“*numerical scales*”). Besides also pictures (“*smiley scales*”) are used incidentally, but these will not be discussed here, because in this case the measurement results are processed in the same way as those obtained using verbal primary scales.

Table 3.1

Examples of items used in time series of happiness in the Netherlands, 1975-2012

| Survey           | Eurobarometer   | World Values Survey   | Statistics Netherlands (CBS)  |
|------------------|---|---|---|
| Item code        | O-HL-c-sq-v-3-ai  | O-HL-u-sq-v-4-a   | O-HP-u-sq-v-5-a   |
| Survey question  | Taking all things together, how would you say things are these days? Would you say you are...?                    | Taking all together, would you say you are ...?   | To what extent do you consider yourself a happy person?   |
| Response options | <ul style="list-style-type: none"> <li>- Really happy</li> <li>- Fairly happy</li> <li>- Not too happy</li> </ul> | <ul style="list-style-type: none"> <li>- Very happy</li> <li>- Quite happy</li> <li>- Not very happy</li> <li>- Not at all happy</li> </ul> | <ul style="list-style-type: none"> <li>- Very happy</li> <li>- Happy</li> <li>- Neither happy nor unhappy</li> <li>- Not very happy</li> <li>- Unhappy</li> </ul> |

### 3/3 Happiness measured as a discrete variable at the ordinal level of measurement

Happiness, when measured as described in Section 3/2, is a discrete variable at the *ordinal level of measurement*, not only if verbal scales are used, but also for numerical scales. The digits of numerical scales are code numbers in principle; cf., e.g., progress report marks in Dutch schools, where usually “6” is defined as a code number for “satisfactory progress”. In Section 6/1.2 we will discuss this into more detail. A variable is called *discrete* if it can adopt a very limited number of values only, four in the example.

Upgrade of the ordinal to the *interval level of measurement* is possible for both verbal and numerical scales, which requires to make additional assumptions (Section 3/6 and 6/2.9).

### 3/4 Happiness research at the individual vs. the collective level

Happiness can be measured at two levels, the individual and the collective one. This distinction has consequences for the way the measured responses are treated in the further analysis.

#### *Individual level*

In studies at the micro level of individuals the researcher links the responses for both happiness and the correlate of the study of each respondent separately. In this way one can investigate whether, e.g., very happy people are living more frequently in a rural than in an urban environment or whether the reverse is true. In this case the study can be continued on the basis of the happiness as it has been measured in Section 3/2. A cross tabulation may provide sufficient information to answer the above research question.

#### *Collective level*

At the macro level of nations this is usually not possible. A scientist who wants to compare the happiness situation of the Dutch population to the French is not interested in all individual responses, but only in the *statistical distribution* of the happiness intensity in both nations separately. 'Measuring happiness at the collective level' is a short-hand term for measuring this statistical distribution of individual responses of the members of this collectivity. The researcher wants to characterize these distributions with one of more *index numbers*. We will use the term "index number" in this context as it is in use among statisticians and economists, i.e. as a number that quantifies a property of a phenomenon in a standard way, so not for the sum score of a number of indicators, as it is used by sociologists.

The next step will be to look for a possible relationship between these index numbers and the index numbers that characterize potentially interesting correlates in both nations. Such correlates do not necessarily bear on individual inhabitants, but can also pertain to living conditions such as, e.g., the climate or the way the government operates in one or more specific respects.

Hence, one of the central issues of happiness research is to obtain useful index numbers that:

- sufficiently characterize the statistical distribution of happiness in a defined population
- on the basis of the measurement of happiness
- in a sample from and considered to be representative for that population
- in the context of a survey
- using a suitable discrete primary scale of measurement

and this is the issue we will focus on in the current contribution.

### 3/5 Cumulative frequency distribution of happiness in a sample

The result of measurement of happiness as described in 3/2 can be summarized completely in a *cumulative frequency distribution*, which is characterized by the “*cumulative distribution function*” (abbreviated as “cdf”) of the measured happiness in a sample. This cdf is a step function with the same number of steps as the number of response options of the primary scale of measurement. We will denote this number with the symbol “ $k$ ”.

In Table 3.2 we give a hypothetical example of a sample with size  $N=200$  and the four-point verbal primary scale ( $k=4$ ) as described in Sec. 3/2.

The *absolute frequencies*  $\{n(j), j=1(1)k\}$ , i.e. the numbers of respondents selecting the same rating  $j$ , are obtained by simply counting. We define the corresponding *relative frequency*  $f(j):= n(j)/N$  and the *cumulative frequency*  $F(j):=F(j-1)+ f(j)$ , with  $F(1):= f(1)$ .

The symbol “ $:=$ ” means “is defined as” and the notation  $j=1(2)7$  means that the variable  $j$  varies from 1 to 7 with steps equal to 2; in this case  $j$  can adopt the values 1, 3, 5 and 7 successively.

Table 4.2  
Happiness frequency distribution in a hypothetical sample

| $j$   | response option<br>(rating) | frequency          |                    |                      |
|-------|-----------------------------|--------------------|--------------------|----------------------|
|       |                             | absolute<br>$n(j)$ | relative<br>$f(j)$ | cumulative<br>$F(j)$ |
| 1     | “not at all happy”          | 20                 | 0.10               | 0.10                 |
| 2     | “not too happy”             | 40                 | 0.20               | 0.30                 |
| 3     | “quite happy”               | 80                 | 0.40               | 0.70                 |
| 4     | “very happy”                | 60                 | 0.30               | 1.00                 |
| TOTAL |                             | $N = 200$          | 1.00               |                      |

Unfortunately the questions used in the various surveys and other investigations are not at all identical, nor are the response options offered. Except the inevitable differences between languages also differences are found in the phrasing (wording) and even in the number of the response options. This strongly hampers the comparison and the application of results obtained in studies in which different primary scales of measurement have been used.

### 3/6 Index numbers for the statistical distribution of happiness

As has been pointed out in Section 3/4 happiness research requires the characterization of the happiness distribution by index numbers, in particular of the distribution of happiness in the population. These index numbers are so-called *statistics*. 'Statistics' in this context are defined as 'variables, the value of which can be calculated from observed data', in our case the sample size, the happiness ratings and their observed frequencies.

Usually two such index numbers are defined. One is for the general happiness level, also referred to as the "*central tendency* of the distribution", for which the *average* or *mean* value is the usual choice; most statisticians make a distinction between "average" for the sample distribution and "mean" for the population distribution. The other index number is to characterize the dispersion, the scatter around the central value, in other words the *inequality* or *disparity* within a distribution. A conventional choice for this statistic is the *standard deviation* of the distribution. The most frequently used symbols for these sample statistics are *m* and *s* respectively.

NOTE: It is conventional to represent variables in scientific documents by symbols printed using italics.

Contrary to a wide-spread belief, not only the standard deviation, but also the average value is to be considered as a statistic that characterizes the happiness *inequality*. Whereas the standard deviation characterizes the inequality *within* a distribution, the average value of the latter has, from a sociological point of view, no other meaning than to characterize the inequality *between* distributions of different nations or other collectivities.

The need expressed above necessitates actions to facilitate the calculation of these statistics. This requires the 'upgrade' of the ordinal response options of the primary scale of measurement into so-called "*cardinal numbers*". Cardinal numbers are numbers that may be subjected to arithmetic operations such as addition, multiplication etc. This transformation process of the *k* primary ratings (labels or digits) into *k* cardinal numbers is referred to as "*cardinalization*" or in this particular case as "*rescaling*". The cardinal numbers obtained in this way are referred to as *secondary ratings*. The various cardinalization methods used in this context are discussed in Section 6/1.

Moreover, the problem of the different numbers (*k*) of primary ratings of scales of measurement has to be solved. The *scale length of a secondary scale* is defined as the difference between the highest and the lowest conceivable value the variable can adopt when using this scale. In the case of an ordinal numbers scale (Sec.6/1.3) these values coincide with the values of the largest c.q. smallest rating of the (secondary) happiness intensity scale; as will be pointed out in Sec. 6/1.5, this does not apply to the case(s) of nonlinear transformation.

Measurement results obtained by using different primary scales can be compared only when at least the corresponding secondary scales have equal lengths; if not, additional operations are required, which will be discussed in Section 6/1.4.

Finally, there is the need to translate the results obtained from the sample measurements into information on the happiness distribution within *the population* that is represented by this sample. The methods to be applied can be distinguished into two mainstreams.

The first one adopts the  $k$ -step discrete cumulative sample distribution on the basis of the secondary ratings as the corresponding distribution in the population as well. These procedures are dealt with in Section 6/2.1.

The other one postulates the existence of a happiness variable in the population, which is (a) unobservable or “*latent*”, (b) *continuous*, i.e. can adopt any real value of the  $[0, 10]$  happiness continuum and (c) characterized by a continuous probability distribution in the population. To obtain the latter, a “family” of suitable distributions has to be specified and then within this family the “member” is selected which fits best to the sample distribution. The methods on the basis of this Continuum Approach, which has been developed by Kalmijn, are discussed in the Sections 6/2.5-6/2.9.

### 3/7 Non-standard methods

In Sections 3/1-3/6, we described how happiness, and more generally subjective well-being, is measured using the standard procedure with the following main characteristics

- direct measurement
- by self-reporting one’s happiness intensity
- answering to a single closed question
- selecting one out of a small number (usually  $\leq 12$ ) unambiguously ordered response options,
- by individuals belonging to a sample that is considered to be representative for the target collectivity of the study.

Besides the standard procedures, there is developing an unstructured and heterogeneous set of less courant methods that deviate from the standard methods in one or more of the above respects. We will refer to these methods as *non-standard methods*. A list of such methods is given in Appendix B.

Examples of these non-standard methods include:

- Self-report using multiple questions, such as:
  - *Affect balance scales*, Such as Bradburn’s [Affect Balance Score](#) (ABS)
  - *Life satisfaction questionnaires*, such as Diener’s [Satisfaction With Life Scale](#) (SWLS)

- Repeated single questions on happiness of the moment, such as:
  - *Experience Sampling Method* (ESM) e.g. Wessman & Ricks [Personal Feeling Scale](#)
  - *Day Reconstruction Method* (DRM), e.g. Veenhoven's [Happiness Diary](#)
- Narrative self-report coded afterwards, such as:
  - *Content analysis of open interviews*, e.g. Mussen's rating of [trait cheerfulness](#)
  - *Content analysis of life reviews*, e.g. Danner's [rating of happiness in autobiographies](#)
  - *Content analysis of diaries*, e.g. Newman Langner [scale of post divorce adaptation](#)
- Estimating happiness by *external judges* on the basis of observation of the subject's behavior, which is assumed to be related to his/her happiness.
  - Rating of cheerful behavior: e.g. Webb's [rating of pupils by teachers](#)
  - Rating on multiple indications of mood, e.g. [wellbeing subscale](#) of Baker and Dowling Interact schedule
  - Rating by interviewer of non-verbal behavior during the interview, e.g. Allensbacher rating of [cheerful appearance](#)
  - Time sampling of happy and unhappy utterances, part of Bailey [observation schedule for babies](#)

These methods require skilled observers, but in some relatively simple cases the observation can be automated, e.g., the frequency and the average duration of crying by babies

Not all non-standard methods can be considered to be a part of the 'body of science', which requires a demonstrated replicability both of the application of the method and of the results obtained in this way. Several of those methods have been developed for one specific application.

On the other hand, not all methods have this pretention. The standard Eurobarometer question is usually inapt to assess the progress of a therapeutic process of a psychotherapist's client, which requires a specific non-standard approach.