6 CONVERSION OF MEASUREMENT RESULTS

Wim Kalmijn[1]

6/1 Cardinalization and harmonization of the ratings of a primary scale

6/2 Methods for estimating the population distribution parameters

6/3 Application of these methods in the WDH

Selected bibliography

6/1 Cardinalization and harmonization of the ratings of a primary scale

6/1.1 Methods to transform the ratings of primary scales

6/1.2 Linear transformation of the primary ratings. Step 1: coding

6/1.3 Linear transformation. Step 2: cardinalization

6/1.4 Linear transformation. Step 3: harmonization

6/1.5 Objections against linear transformation. Thurstone ratings

6/1.6 The Scale Interval Study

6/1.1 Methods to transform the ratings of primary scales

Scale transformation or "*rescaling*" is defined in Section 3/6 as the operation in which all *ratings* of the primary ordinal scale are replaced with appropriate cardinal numbers. This cardinalization is necessary to enable one to calculate the desired statistics such as, e.g., the average happiness value or the standard deviation of the

[1] Text June 2015

sample. Moreover, the problem of the unequal ‘scale lengths’ (Section 3/6) has to be solved. This is done by transforming the ratings onto a secondary scale bounded with scale values that are common for all secondary scales; for these boundaries the values “0” and “10” are selected usually, corresponding to the situation that characterizes the least and the most happy situation respectively. This operation is referred to as “*scale harmonization*”; see Section 6/1.4.

Veenhoven has developed three methods in the course of the years for application to verbal scales:

- linear transformation of the ordinal numbers (ranks) after cardinalization [Section 6/1.2 – 6/1.4],
- transformation into a “fixed” happiness value for each textual label separately on the basis of semantic judgment and to apply these values as general secondary ratings, which are also referred to as “Thurstone ratings” [Section 6/1.5] and
- transformation into “mid-interval values” on the basis of judgment of the response options in the context of the total primary scale [Section 6/1.6].

Both the second and the third transformation method above, have been developed specifically to be applied to verbal and pictorial scales. In the case of numerical scales, linear transformation is the standard option.

6/1.2 Linear transformation of the primary ratings: *Step 1: Coding*

This transformation process proceeds in a number of steps with distinct objectives; strictly speaking, only the third transformation step (6/1.4) can be characterized as “linear”.

In the case of *verbal scales* the first step of the transformation process is “*coding*”, i.e. the conversion of each alphanumerical “*label*” into a corresponding code. This step is made during the processing of the response forms. The usual coding system assigns to each response option the ordinal number of the order of presentation. If the most happy option is presented in the survey as the first one, the order is sometimes, but not always reversed to assign the digit “1” to the least happy option. For *pictorial scales* a similar procedure is applicable.

If the primary scale is a *numerical scale*, then the response options are ordinal numbers already, so this coding step can be skipped.

Incidentally scales are presented to the respondents without any verbal, pictorial or numerical label, except for both extreme so-called *anchors*, e.g., “extremely unhappy” and “extremely happy” respectively in the example:

□ □ □ □ □ □ □ □ □ □ □

extremely unhappy   extremely happy.

Here the ticked box is replaced afterwards with its ordinal number by the surveying institute. Such scales are treated in the same way as numerical, i.c. 11-point primary scales.

6/1.3 Linear transformation: *Step 2: Cardinalization*

The next step is that the researcher 'upgrades' the ordinal numbers of the verbal ratings to cardinal numbers. This transformation is done by simply replacing the ordinal with the corresponding cardinal numbers. If necessary, the scale is "reversed" to ensure that the least happy rating always has the value "1".

In this way the researcher introduces equidistance of pairs of consecutive ratings on the scale, if not unconsciously, at least unintentionally, without any justification, in particular when the primary scale is a verbal one.

6/1.4 Linear transformation: *Step 3: Harmonization*

Measurement results obtained by using different primary scales can be compared only when at least the secondary scales have equal lengths; if not, they have to be harmonized. "*Harmonized scales*" are defined in this context as scales in which the happiness variable ranges between *common* boundaries. For these common boundaries the values "0" and "10" are the usual choice, corresponding to the least and the most happy conceivable situation respectively.

The scale harmonization in the case of linear transformation is achieved by stretching or compression of the scale as obtained in Section 6/1.3 to a *secondary scale* in such a way that to the least happy response option the secondary rating "0" is assigned and to the most happy the rating "10". The other secondary ratings are equally spaced in between. All secondary ratings are cardinal numbers. The harmonization method in this case is also referred to as "*linear stretching*" or as "*direct rescaling*".

A widely spread misunderstanding is that linear transformation always results in a secondary 11-point scale. However, in linear transformation of e.g. a 4-point scale all four *ratings* are transformed, but no new ratings are introduced:

Table 6/1
Transformation of verbal response options

Primary response option →	code (order) →	rating before →	after stretching
“very happy”	“4”	4.00	10.00
“quite happy”	“3”	3.00	6.67
“not too happy”	“2”	2.00	3.33
“not at all happy”	“1”	1.00	0.00

A consequence is that a primary numerical 10-point scale {1, 2, …, 10} is stretched by linear transformation into a secondary 10-point scale, but with a distance 10/9 between successive ratings: {0, 10/9, 2x10/9, …, 8x10/9, 10}. Scales with more than 11 ratings are compressed, but such scales are rather exceptional.

The general formula for the linear transformation of the rating *j* (1≤*j*≤*k*) on a primary *k*-point scale {1,2,…,*k*} into a secondary rating on the 0-10 scale is:

$$j \longrightarrow \frac{10}{k-1} \cdot (j-1)$$

This transformation is called “linear” because the ‘new’ rating is a linear function of the ‘old’ *j*, where for all *k* ratings both the coefficient and the constant term have the same values, in casu 10/(*k*–1), and –10/(*k*–1) respectively.

The *discrete 11-point numerical 0 to 10 scale* has a particular position within the class of numerical scales. All ratings of other primary numerical are ordinal numbers, but “0” in this particular scale is not an ordinal number. After cardinalization, the {0, 1, 2, … , 10} scale is harmonized automatically and linear stretch has no additional effect at all. This property has probably made this scale increasingly the favorite.

6/1.5 Objections against linear transformation: *Thurstone ratings*

Several methodological objections can be raised contra the procedures described under 6/1.3 and 6/1.4, in particular when they are applied to verbal scales.

The most fundamental is that in no way one allows for the semantics of the labels: the wordings of the response options on the primary scale are completely ignored by the researcher, although they are essential for the respondent’s choice when compared to the presented alternative ratings. Moreover, the assumption of equidistance both before and after stretching is most speculative, at least for scales with few options, with no other justification than the “simplex sigillum veri” principle in a situation of complete ignorance on this.

To meet these objections, in particular the first one, the so-called “Semantic Judgement of Fixed Word Value” has been introduced. Full weight in this approach is given to the wording of the primary response options. A number of experts in this research area have agreed in 1993 that, e.g., the option “very happy” or its translation is transformed directly into the fixed secondary rating 9.3 on a 0-10 continuum *in all situations*, irrespective of the number, the scale position, the labels of alternative options and their language. In this way one also gets rid of the equidistance artifact.

The principle of this has been proposed by Thurstone, and hence the secondary ratings obtained in this way are reported in the WDH as “*Thurstone ratings*”.

Contrary to linear transformations of the type $x \rightarrow y = a + bx$, such as the transformation $j \rightarrow 10(j–1)/(k–1)$ in Sec, 6/1.4, the direct transformation of the primary into the corresponding secondary rating (e.g. “very happy” → 9.3) is a *nonlinear* transformation.

In such situations, the terminal secondary ratings do not coincide with the boundaries of the common continuum of conceivable happiness intensities, usually chosen at the values 0 and 10 respectively, and the maximum distance between two secondary ratings will be less than the scale length (10). A respondent may judge his own happiness as, e.g., about 9.7, but he has to report the closest rating (“very happy”), which is processed in the survey analysis as 9.3.

The concept of a fixed value of the secondary rating assigned to a specific label, however, is not satisfactory in all respects. Both the wording of alternative options and the position on the scale with respect to these alternatives are completely ignored. The same holds for linguistic and other cultural differences between nations.

The introduction of an alternative approach, which is described in Section 6/1.6 together with further developments (Section 6/2) has made the application of Thurstone ratings in practice almost completely obsolete.

6/1.6 Semantic judgment of contextual value of the ratings

The above objections against the Thurstone ratings gave rise to the start of the *International Happiness Scale Interval Study*

The fundament of this study was the recognition that, e.g., the response option “pretty happy” does not cover one single happiness intensity only, but that respondents may also assign the same qualification to intensities in its immediate environment. In other words: each verbal response option corresponds to a number of successive happiness intensities. The total range of the 0-10 scale is a continuum that can be partitioned into k disjunct subintervals, each of which being linked to one of the k response options of the primary scale. Every subinterval has an upper and a lower boundary. Each upper boundary coincides with the lower one of the adjacent interval, except for the "highest" one. Two of the *k*+1 boundaries are fixed at the vales "0" and "10" respectively, the other *k*–1 can be determined empirically by asking a number of selected 'judges' to express one's opinion on the position on a [0, 10] continuum of the boundary between two adjacent options, e.g., the boundary between "pretty happy" and "very happy" on this *specific scale*, i.e. within this specific question, with these specific *k*–1 alternative response options and in this specific language. The semantic judgment of the labels like "very happy" in this way is always made in the *context* of the total happiness measure.

Instead of "boundary" also the terms "*transition points*", "*cutting points*" and "*cut points*" are in use. Incidentally they are also referred to as "*threshold values*", but this may give rise to some confusion: in medical-biological research the term "threshold values" is generally indicating a lower limit, whereas the usual numbering in our case refers to the upper boundary of the subintervals. In the case of a *k*-point scale the symbol b_k =10 indicates the upper boundary of the scale and b_o=0 the lower boundary of the [0, 10] continuum.

The judges in the above example have been selected on the basis of their mother tongue being English. In the case of a primary scale with French labels, the mother tongue of the judges should be French. The underlying assumption is that the words "heureux" (fr), "happy" (en), "gelukkig" (nl) etc. do not necessarily express exactly the same experience.

In practice, these studies are run computer-based, using an instrument, which has been devised particularly for this application and is called the *Scale Interval Recorder*.

Judges have to establish the positions of all *k*–1 nontrivial boundaries by shifting the sliders towards the appropriate position and fixing the latter.
The average position, averaged over all judges, is adopted as the position of this particular boundary in this particular case. A case in the context of the Scale Interval Study is defined as the collection of the average positions of all boundaries within one specific combination of a measure of happiness (or "item") and nation/language within a short period. A "study" or "session", as defined within this specific context, consists of a number of cases (up to 10), in principle all treated subsequently by the same group of judges.

In the case of this scale transformation, the position of the middle of the subinterval corresponding to the option "pretty happy" is referred to as its *mid-interval value* (abbreviated as MIV) and is adopted as its secondary rating, albeit within this particular measure of happiness only. In the same way as the Thurstone transformation in Section 6/1.5, the direct transformation of primary verbal ratings into the corresponding MIV is nonlinear.

A survey of the various transformation methods is given in Scheme 6/1

Scheme 6/1
Transformation of primary into corresponding secondary ratings

	PRIMARY RATING SCALES (direct measurement)	(transformed) SECONDARY RATING SCALES not harmonized	 harmonized
level of measurement number of ratings scale type of ratings	ordinal k (usually $3 \leq k \leq 12$) discrete ciphers or tekst	interval k discrete cardinal numbers	interval k discrete cardinal numbers
ratings	ciphers **[numerical scales]** $\{1, 2, ..., k\}$ or textual labels **[verbal scales]** e.g. with k =3: {"not too happy", "fairly happy", "very happy"}	first k natural numbers≠0 $(1, 2, ..., k)$ linear transformation or linear stretch nonlinear transformation	k real numbers within the [0; 10] interval (a) k equidistant numbers $\{0, 10/(k\text{-}1), ..., 10\}$ (b) fixed semantic e.g. Thurstone ratings (c) contextual semantic (midinterval values)
Identical calculated sample statistics and corresponding estimated population statistics		'untransformed' mean and standard deviation	'transformed' mean and standard deviation

6/2 Methods for estimating the population distribution statistics

---

6/2.1 Sample statistics

6/2.2 Happiness distribution in the sample and in the population

6/2.3 Latent versus measured happiness

6/2.4 The underlying measurement model

6/2.5 Cumulative distribution of happiness within the population

6/2.6 Index numbers of the estimated happiness distribution in the population

6/2.7 The conversion procedure

6/2.8 Models for the population happiness distribution

6/2.9 Application of the continuum approach to numerical scales

6/2.10 Indirect calibration of verbal scales according to the Reference Distribution Method

---

6/2.1 Sample statistics

The ultimate objective of measurement of happiness is to obtain information on the happiness distribution under investigation, more specifically to obtain estimates of its main index numbers: its mean value and its standard deviation.

As has been pointed out in Section 3/6, the methods to convert the sample measurement results into information on the happiness distribution within *the population* that is represented by this sample, also referred to as the "target population", can be distinguished into two mainstreams.
In this section we describe the methods of the first one; those of the second will be discussed in Section 6/2.6- 6/2.10.

All methods in the first mainstream follow the same basic pattern, which can be summarized as follows:

(1) Measure the happiness distribution in the sample from the target population in terms of primary ratings and corresponding relative frequencies (Section 4.5)
(2) Transform the primary ratings into the corresponding secondary ratings, either through linear transformation (Section 6/1.2 – 6.1/4) or through nonlinear transformation (Section6/1.6)
(3) [Optional in the case of linear transformation only] Harmonize the secondary rating scale.

(4) Calculate the sample average value as the weighted sum of all secondary ratings, using the corresponding relative frequencies as weights.
Calculate the standard deviation within the sample in the traditional way on the basis of the sample average value, the secondary ratings and the relative frequencies

(5) Adopt the values of the statistics obtained under (4) as the estimates of the corresponding statistics of the happiness distribution in the population.
The estimated mean value is sometimes referred to as the “*happiness score*” of this population.

Ad (2) The transformations described in Section 6/1 are necessary to enable the characterization of the happiness distribution in the sample by appropriate index numbers. Therefore, the transformation is followed by the calculation of two sample statistics: the average value and the standard deviation of the distribution obtained in this way.
The sample average value *m* is calculated as the weighted average value of all secondary ratings, with the relative frequency as the weighing factor of the corresponding rating. In the case of a *k*-point scale, *m* is obtained as

$$m = \sum_{j=1}^{k} f_j r_j$$

where f_j **:**= the relative frequency of the secondary rating r_j (j=1, 2, …, k). The standard deviation within the sample is usually denoted “*s*” and is obtained as

$$s = \sqrt{\sum_{j=1}^{k} f_j (r_j - m)^2}$$

Ad (3) A particular situation occurs when the primary ratings are transformed directly into their ordinal numbers ($r_j = j$). Application of the formulae given above under ad (2) results in statistics which are sometimes recorded in the WDH as “*untransformed statistics*”. This qualification, however, is a misleading short-hand expression for “*incompletely transformed*” , since both the transformation of the response options into the ordinal numbers and the transformation of the ordinal into the cardinal numbers have been realized. Only the harmonization (Section 6/1.4.) is still missing, so the correct name would be “*unharmonized*”. The statistics previously referred to as “*transformed statistics*” are also *harmonized*.

Harmonization by linear transformation is an option only to be applied to cardinalized ordinal numbers. In the case of nonlinear transformation the secondary scale length is always introduced as its common value (usually 10) and harmonization is built-in already.

Ad (4) In the case of linear transformation, an alternative way to compute the harmonized statistics {*mh*, *sh*} is the linear transformation of the corresponding unharmonized statistics {*mu*, *su*} :

$$mu \longrightarrow mh = \frac{10}{k-1} \cdot (mu - 1)$$

and

$$su \longrightarrow sh = \frac{10}{k-1} \cdot su$$

Statistics *m* en *s*, obtained in this way have a very limited value, even when they are harmonized. Their comparison is admissible only in studies in which happiness is measured using the same measure of happiness *in all cases* within that study. Nevertheless their application is standard in the vast majority of all happiness studies, in particular the older ones, say before 2010.

6/2.2 Happiness distribution in the sample and in the population

The eventual intention of all this is not to obtain information about the sample, but about the corresponding population. In this final step the process is often derailed.

Many researchers consider the population as a large sample, even the largest possible one. On the basis of that idea their sample results, including the values of the statistics are simply generalized and declared to apply to the target population as well. The average happiness *m* in the sample is adopted as an unbiased estimator of the mean happiness in the population. In the same way, the population variance is an unbiased estimator of the population variance. If the sample size is not too small, the standard deviation within the sample is an almost unbiased estimator of the population standard deviation; for *N* >100, this bias, being <0.5%, can be ignored in practice.

If, however, happiness is measured using a 4-step verbal scale, then due to this choice the sample consists of four kinds of respondents, e.g., very happy people, pretty happy people, not too happy people and people who are not at all happy. In the generalizing approach adopted above, the population also consists of four kinds of people as regards their happiness. These are exactly the same kinds in more or less the same proportions. A second researcher who investigates the happiness situation in the same nation during the same period, asking the same question, but who has a personal preference for a 3-point scale, will eventually conclude to the existence of three kinds of people, which, as a matter of fact, are different kinds as compared to those of his colleague. Which one is the ‘true’ happiness situation in that nation?

It is unsatisfactory that this is fully dependent on a personal preference of the researcher involved. Even more unsatisfactory is that this problem seems hardly to be felt to be an issue.

These comments apply to all rescaling methods, including the first version of the scale interval method, the initial goal of which was to deliver MIV's as an improved alternative to the Thurstone ratings.

6/2.3 Measured versus latent happiness

The problems described in Section 6/2.2 can be solved only by making a consistent distinction between happiness measured in the sample and the *latent happiness* in the population.

Because latent happiness is unobservable by definition, it has to be based on one or more assumptions. It is obvious to choose a latent variable that can adopt, at least in theory, any real value between 0 and 10 including both these limits, in other words a *continuous variable*. We shall denote this variable as *H*; the more happy an individual is, the larger his *H*-value. The variable *H* is a "*random variable*", i.e. it has a probability distribution within the population. Since the variable is unobservable, its probability distribution is unobservable as well, so its type has to be postulated by the researcher. The index numbers that characterize this distribution, e.g., the mean value, are essential in happiness research, because all investigations of the association of happiness with time or other correlates should be based on (estimates of) these index numbers.

Schematically

variable	measured happiness	latent happiness *H*
character	observable	assumed
distribution	frequency distribution	probability distribution
obtained by	measuring (counting)	estimation
in	sample	population
domain	discrete {*k* values}	continuous on [0, 10]
application	estimation of latent happiness distribution	quantification of trends and other associations
link	the measurement model	

6/2.4 The underlying measurement model

Estimation of the index numbers of the happiness distribution in the population requires the existence of a relationship between the measured happiness in the sample and the latent happiness in the population to be known. This relationship is specified in what is called the "*measurement model*". Below we specify such a measurement model, which links up to the approach of the Happiness Scale Interval Study (Section 6.1.6). A schematic representation is given below:

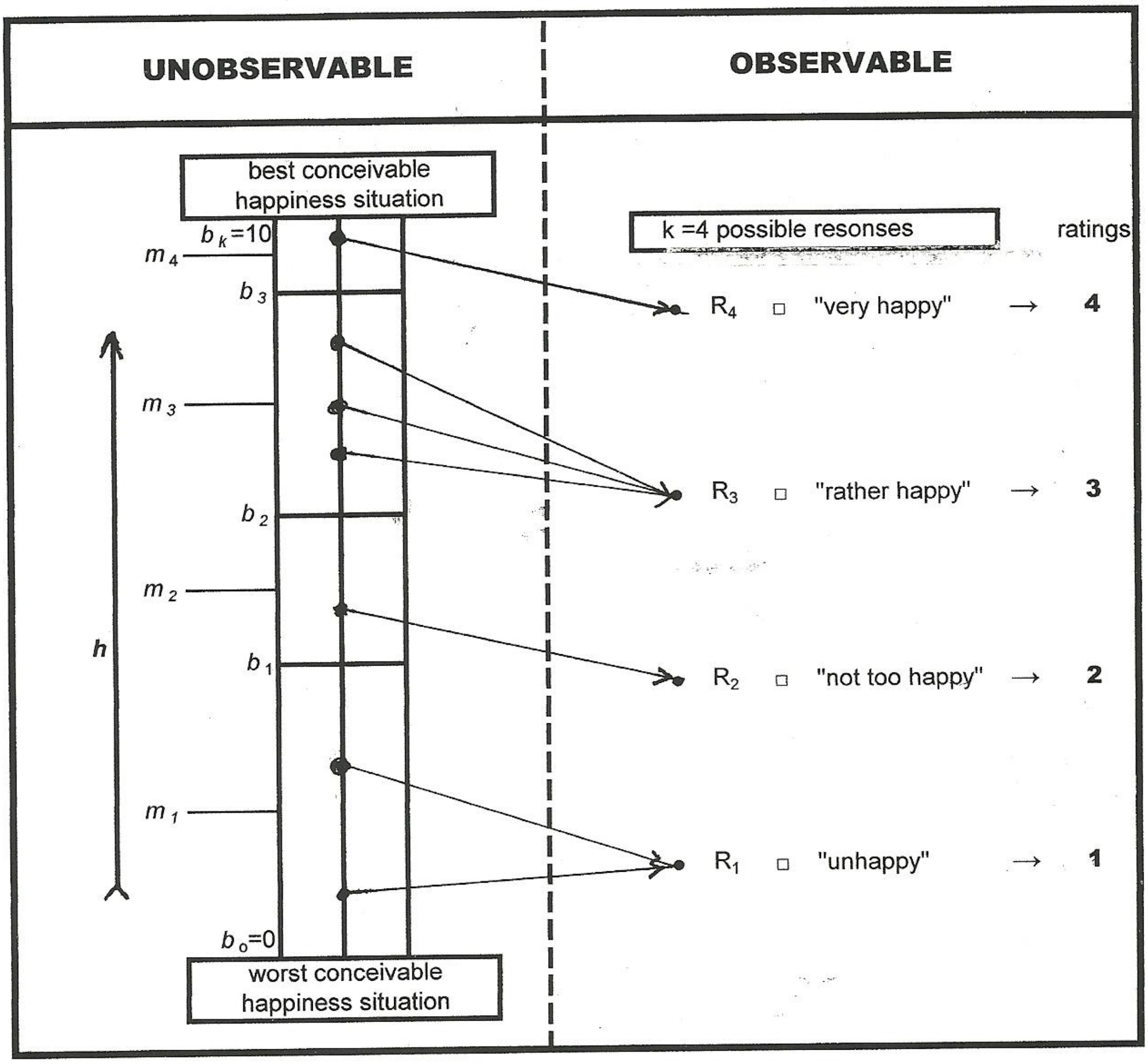


Figure 1 Representation of model for International Scale Interval Study

The continuum [0, 10] is partitioned into a number of subintervals in such a way that

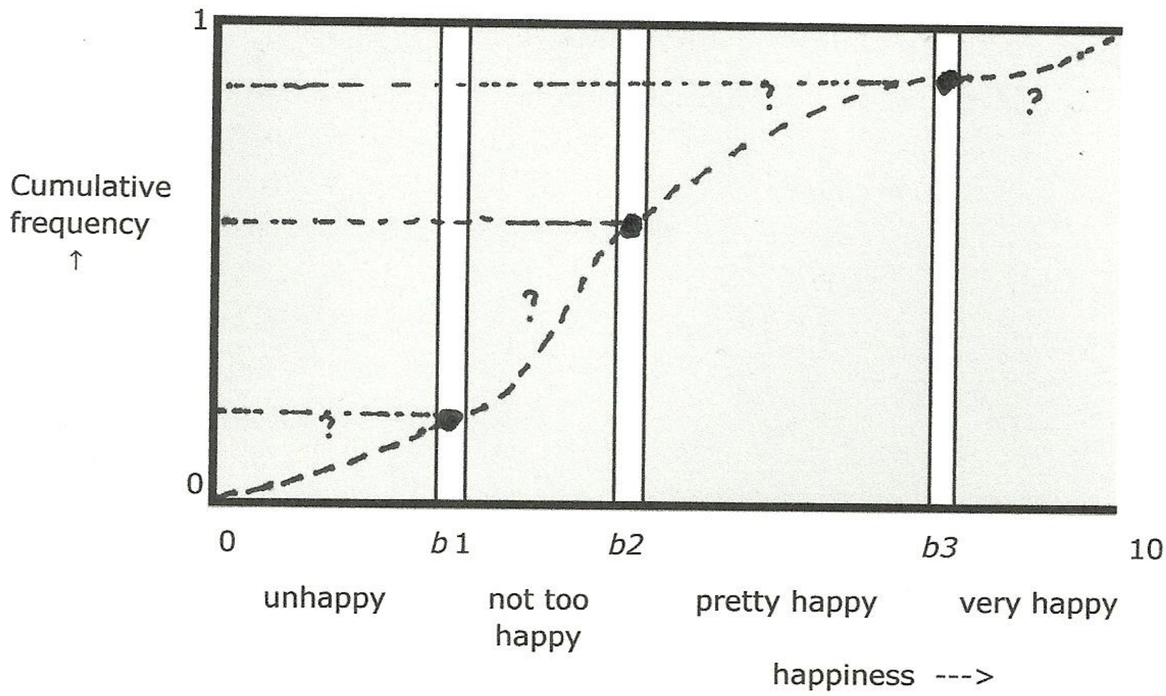
- the number of subintervals is equal to the number of response options *k*
- each interval is linked to one and only one response option
- an interval with larger *H*-values is linked to a response option that describes a happier experience situation

— within the population a reasonable consensus exists on the *k*-1 *H*-values of the boundaries between the *k* subintervals, given the survey happiness question and the presented response options This implies that the positions of all *k*-1 boundaries within the [0, 10] continuum can be estimated
— all subjects within the population with *H*-values belonging to the same subinterval will select the same response if they are included in the sample; this response is the one linked to 'their' sub-interval.

6/2.5 Cumulative distribution of happiness within the population

If *h* represents a given value of the latent happiness variable *H*, the *cumulative distribution function* (cdf) is written as *G*(*h*) and is defined as the relative frequency in the population of *H*-values that are less than or equal to that value *h;* therefore *G* is a monotonous non-decreasing function of *h* and 0 ≤ *G*(*h*) ≤ *G*(10)=1.

This distribution is not observable as a whole, but there are *k* values of the latent variable *H* where *G* is estimable. These are the *H*-values of the *k*–1 non-terminal boundaries of the subintervals and the trivial case *H*=10 where *G*(10)=1.
At the first boundary *G* is the relative frequency of the least happy response option. In each successive boundary value the relative frequency of the next option is added.



The situation is as if the cdf curve is hidden by a screen, with at *k*-1 places, at the boundaries of the subintervals of the response categories, a vertical narrow chink where one 'point' of the curve is observable. However, what one sees is an estimate, because there is some inaccuracy, both horizontally and vertically. The first is caused by the fact that the position of the chinks is determined by the judges and (i) there is no perfect consensus among the judges, nor among the respondents, and (ii) the number of judges is finite. The vertical inaccuracy is caused by the similar fact if the number of respondents is less than the total population size.

6/2.6 Index numbers of the estimated happiness distribution in the population.

Eventually, happiness researchers are not interested in these *k* points of the distribution, which are based on arbitrary choices of the institute that organizes the survey (Sec 3/2). If someone else raises the same question, but presents different response options, he will estimate different points. These points, however, are – in principle – points of the same cdf curve, provided the measurement results are delivered by samples of more or less the same population with a moderate time lag if any. This is a strong point of this method and we will pay more attention to this in Section6/2.10.

Happiness researchers are – or at least should be – interested in the cdf itself, more precisely, its parameters, which are estimates of the population mean and the standard deviation or enable one to estimate these index numbers. These need these index numbers for the investigation of the association of happiness with time and other correlates.

These parameters, however, are dependent on the distribution model that is chosen by the researcher. Nobody knows how the cdf looks like between the chinks, so we have to make assumptions on the model that describes this. One such a possible choice is the "*semi-continuous model*", in which all *k* successive observed points are connected by straight line segments, resulting in a cdf that is represented by a broken line with a kink in each transition point.

6/2.7 Conversion procedure

The conversion of the sample measurement data into the estimated characteristics of the population happiness distribution is the result of four different contributions: See scheme 6/2.7.

The steps in this process constitute a *conversion procedure*. The sample distribution frequencies are – together with the other contributions – converted into characteristics of the population distribution. It would be incorrect to use the term "transformation", since the latter pertains ratings. Transformation describes a process starting with primary ratings, but always ending up with an equal number of secondary ratings, albeit different ones; the whole procedure is confined to the sample situation. In case of the continuum approach, however, no secondary ratings are involved. The population distribution as described above, has no ratings at all, only model parameters and distribution characteristics, so term "transformation" ought to be avoided for linking both distributions.

Scheme 6/2.7
Steps in the conversion procedure

step	*contribution*	*contributor(s)*
1.	choice of "happiness measure" (leading question + response options)	survey conductor
2.	individual happiness frequency distribution (application phase)	respondents jointly
3.	estimation of boundary positions (calibration phase)	judges jointly
4.	choice of the distribution model of the latent population happiness variable	methodologist

6/2.8 Models for the population happiness distribution

For the choice in the fourth contribution in the previous section two alternative models have been developed: a semi-continuous model and a fully continuous one.

In the *semi-continuous model* (Section 6/2.6), the domain of the latent happiness variable, usually [0, 10], is partitioned into k intervals, each corresponding to one of the k response options. A uniform distribution of the latent happiness variable is postulated to exist between each pair of consecutive interval boundaries. The probability density function is not continuous, but it is a step function with a step at each transition value of the latent variable. It can be proven that the estimated mean happiness value of this model is equal to the estimated sample average value as it is obtained by the re-scaling of the observational data on the basis of the mid-interval values according to Veenhoven (Section6/1.6). This value is the weighted average of the MIV, each weighted with its observed frequency in the sample. An unbiased estimator of the population variance can be obtained as the sample variance s^2, augmented with a term accounting for the fact that respondents of the same option may have unequal H-values within the same subinterval.

In the *fully continuous model* the distribution of the latent happiness is postulated to be the beta distribution with [0, 10] as its domain. The beta distribution is a two-sided bounded distribution with two positive parameters, usually referred to as its *shape parameters* and denoted as α and β, covering a wide class of distributions. For $\alpha > 1$ and $\beta > 1$, the density is unimodal and equals zero at both domain boundaries. The values of these parameters are estimated from the observations according to a method which is known as the "maximum likelihood estimation method". On the basis of the relationships between these parameters and the mean value and the variance of the population distribution, the latter characteristics can be estimated.

It is recommended to pay attention to at least the four points below when making a choice in favor of the most suitable probability distribution model for happiness in a population:

(a) In the fully continuous model both the cdf $G(h)$ and the probability density $g(h)$ are represented by smooth curves. This is not the case in the semi-continuous model, where the cdf is represented by a broken line with k-1 kinks at the H-values that have been reported as the boundary values by the judges in a scale interval study. The probability density is a step function of h with a step at each of these transition points. The number of steps follows from the choice of the number of response options for the sample. In the case of the fully continuous model, however, the subjective choice of the number of response options does not influence the shape of $G(h)$ and $g(h)$, which is an argument to prefer the latter model.

(b) The precision at which the mean value of the semi-continuous distribution is estimated can be calculated. This enables one to construct an e.g. 95 % confidence interval for the true but unknown mean happiness value in the population. In the case of the fully continuous method, this is not (yet) possible, so we do not know then the precise quality of these estimates. This might be an argument to prefer the semi-continuous model.

(c) It should be borne in mind that in principle different models give rise to different values of the estimated population mean value. The cause of this phenomenon is in the definition of the mean value and of the standard deviation in the population, which contain the probability density $g(h)$, being defined in a different way for different models.

Generally speaking, the estimated population mean value will be closer to the value h=5 if the semi-continuous distribution model is adopted than in the case of a choice in favor of the fully continuous model, albeit that the differences are modest. Moreover, application of the semi-continuous model gives rise to usually (slightly) larger estimated standard deviations than the fully continuous model.

An unjustified recommendation would be to choose in favor of the model that delivers the most attractive estimates and is expected to have a better

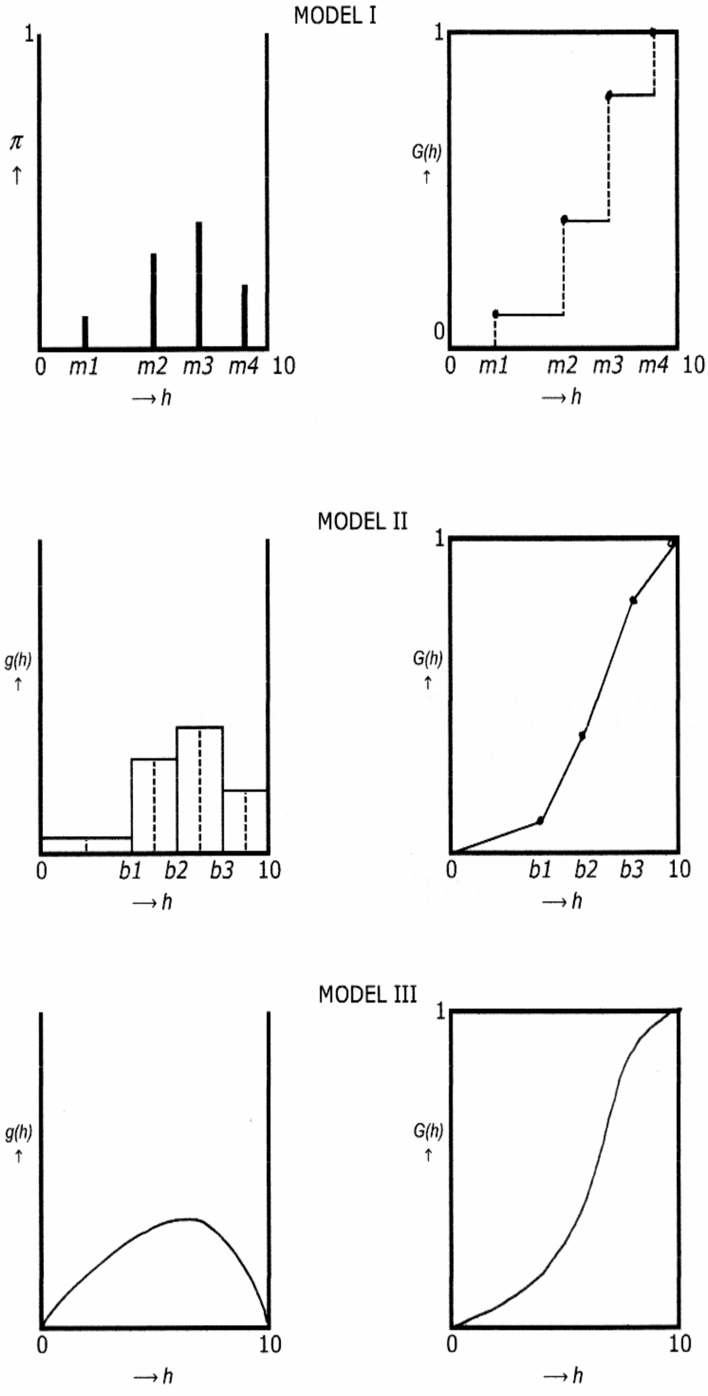


Fig 3 Probability values and densities (left) and cumulative probabilities (right) for $h \in [0, 10]$ in three models: I (discrete distribution), II (semi-continuous distribution) and III (beta distribution), all on the basis of a four-point rating scale.

discriminating power; usually this is the fully continuous model. The conclusion that the choice of a 'better' model would give rise to happier people and/or less inequality in a nation is a misconception. Comparison of nations on the basis of different models is not admissible for similar reasons.

(d) The semi-continuous model has always a perfect internal "goodness-of-fit", contrary to the fully continuous model in the case k>3. use of this difference is in the number of model parameters of both models. In the semi-continuous model 2k–2 independent parameters are estimated, k–1 for the h-values of the transition points and k–1 more for the different values of the probability density. The fully continuous model on the basis of the beta distribution, however, has two independent parameters only (α and β). Apart from the question to what extent a goodness-of-fit test is meaningful in situations like these, the conclusion should be that any comparison on the basis of different models in this respect would be incorrect from scientific point of view.

Taking all together, the provisional choice to be applied in the WDH has been made in favor of the full continuous distribution on the basis of the beta distribution.

6/2.9 Application of the continuum approach to numerical scales

The continuum approach has been developed for application to verbal scales, but can be used, mutatis mutandis, equally well in the case of pictorial scales.

Besides, the principle is also applicable to numerical scales. Only will the third contribution in the scheme in Section 6/2.7 not be delivered by 'judges', but by the methodologist. The [0, 10] interval will be partitioned into k subintervals in the case of a k-point numerical primary scale. As long as no tractable alternative has been proven to be more correct, these subintervals ate given equal widths 10/k.

Now each primary rating is considered to be the mid-interval value of the corresponding subinterval. The postulated equidistance is judged to be acceptable for not too small values k-values. In the WDH, the arbitrary value k = 7 has been adopted as the smallest acceptable number of primary numerical ratings.

The sample statistics m and s of the primary rating scales of the type {1, 2, … , k} can be obtained as is described in Section 6/2.1. Conversion of these statistics yields unbiased estimates of the corresponding index numbers of the population happiness distribution as follows for:

the population mean $(10/k)(m–½)$,
and for
the population standard deviation $(10/k)\sqrt{(s^2+1/12)}$,

so in the case of a discrete 10-point scale {1, 2, …, 10} these estimates are m–½ and $\sqrt{(s^2+1/12)}$ respectively.

The above mentioned numerical scales are all based on cardinalized ordinal numbers. Sometimes, however, other discrete scales are used, e.g., with the value ”0” as its lowest primary rating. In practice this is restricted almost completely to the 11-point scale {0, 1, 2, …, 10}. In this case k=11 and the conversion formulae are for

the population mean $(10/11)(m+½)$,

and for

the population standard deviation $(10/11)\sqrt{(s^2+1/12)}$.

Note the “+” sign in the formula for the mean in this case !

If all *primary* ratings are considered as mid-interval values of intervals with equal width, the terminal intervals of the primary {0, 1, …, 10} scale cover the ranges [–½, +½] and [9½, 10½] respectively, i.e. [–½, 10½] for the total primary scale. For the *secondary* 0-10 scale the 11 intervals are compressed; the two terminal intervals are [0, 10/11] and [9+1/11, 10] respectively. It is essential that this distinction is made clearly in the reporting of the statistics. This is done by assigning them different codes (Section 6/4.2).

The proofs of these formulae are given in Appendix A. They are all based on the application of the *semi-continuous model* (Section 6/2.8).

Conversion on the basis of the *fully continuous mode*l with the beta distribution is also possible, in a similar the same way as is described in Section 6/2.8 for verbal primary scales, but adopting 10j/k this time as the position of upper boundary of the j-th subinterval, k being the number of ratings of the primary scale.

However, while for conversion on the basis of the semi-continuous model, knowledge of the sample distribution statistics m and s only is sufficient, conversion on the basis of the fully continuous model requires knowledge of the complete frequency distribution of the sample.

6/2.10 Indirect calibration of verbal scales according to the Reference Distribution Method.

In Section6/1.6, we have described how verbal scales can be calibrated by partitioning the [0, 10] continuum into k subintervals. The positions of the boundaries between these subintervals were found by asking a number of autochthonic ‘judges’ for their opinion on these.

Under favorable conditions verbal scales can be calibrated in a different way, according to the ”Reference Distribution Method”, which has been introduced by DeJonge. We will illustrate the principle using an example:

In the European Social Survey of 2008 in the Netherlands (further abbreviated ESS2008NL) an 11-point scale {0, 1, 2, … , 10} with measure code O-HL-u-sq-n-11-a has been used. The leading question was “Taking all things together, how happy would you say you are?” The best fitting beta distribution was estimated as described in Section 6/2.9 and the parameter estimates are found to be α=10.4 and β=3.4. This distribution has a mean value 7.55 and a standard deviation 1.12.

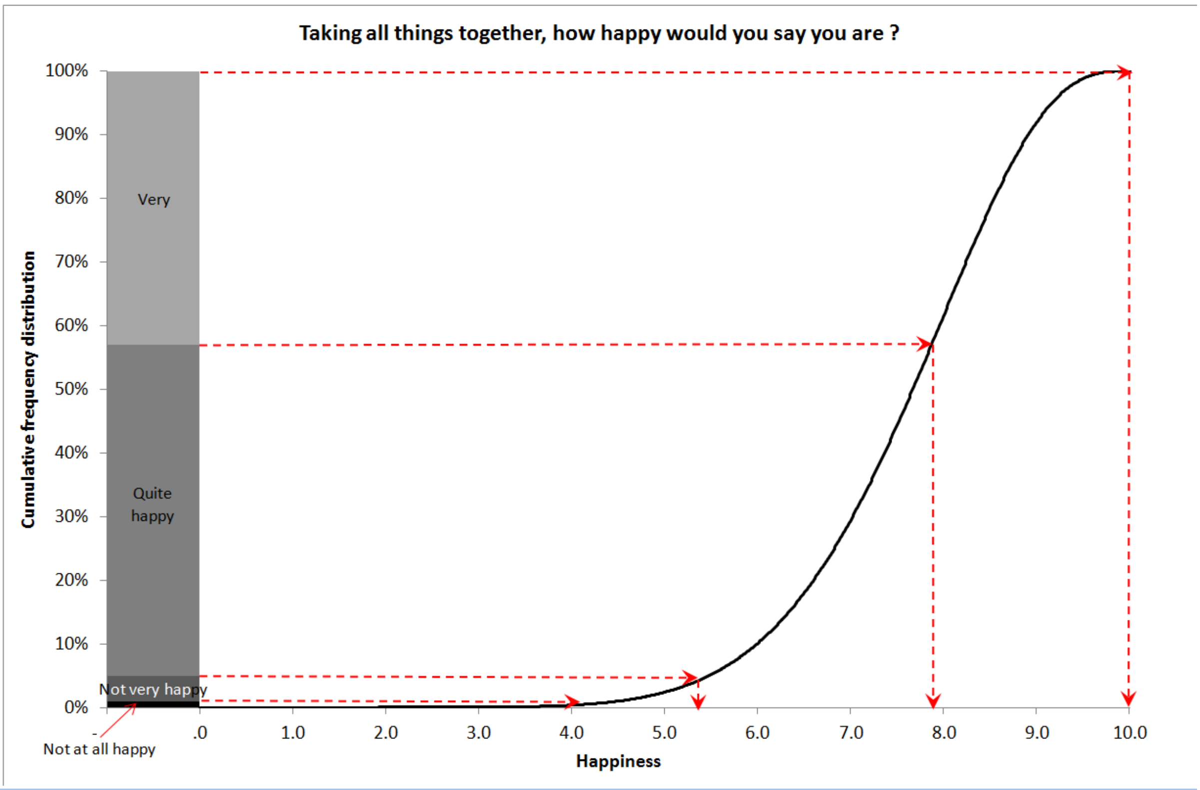
In the about same period (2006) the same leading question was also used in a survey by Eurobarometer, in which a verbal scale is used with the response options {“very happy”, “quite happy”, “not very happy”, “not at all happy”}. The WDH-code of this measure is O-HL-u-sq-v-4-a.
The observed relative response frequencies in this survey were 0.01 (“not very happy”), 0.04, 0.52 and 0.43(“very happy”) respectively, so the cumulative frequency of “quite happy” amounts up to 0.01+0.04+0.52=0.57.

The basic ideas behind this indirect calibration are:

(1) If one asks the same question to samples from the same population in more or less the same period, one estimates different points, but these belong to about the same cdf curves, since the latent happiness distribution has been introduced as independent of the scale of measurement.
(2) If one adopts the cdf obtained using the 11-point scale as the *reference distribution* in this situation, the verbal scale can be calibrated by the comparison of the application results to the reference distribution.

In our example the boundary between “very happy” and “quite happy” corresponds to a cumulative frequency 0.57 (see Figure 6/2.10).
In the cdf of the ESS2008NL study a cumulative frequency of 0.57 can be calculated to correspond to an *H*-value of 7.86 on a [0, 10] continuum. Ergo this is the estimate of the position of the upper boundary of “quite happy”. In this way the other boundaries can also be estimated without the mobilization of judges.

Figure. 6.2.10
Example of the application of the Reference Distribution Method



6/3 Application of these methods in the World Database of Happiness

---

6/3.1 Policy with respect to report measurement of SWB in the WDH

6/3.2. Coding of the index numbers in the WDH

6/3.1 Policy with respect to report measurement of SWB in the WDH

1. In the “Happiness in Nations” component of the WDH, the estimated population mean and standard deviation of the distribution on the [0, 10] continuum of the latent happiness variable in the population of any relevant subjective well-being (SWB) study are reported whenever these statistics are available.

2. Whenever possible, the fully continuous model on the basis of the beta distribution will be adopted. If, however, only the average value and the standard deviation of the sample distribution are known, the semi-continuous model is the alternative, provided the scale of measurement is a numerical scale with at least seven ratings.

3. Whenever necessary and possible, results of previous surveys are converted into findings that are in agreement with the continuum approach.

4. When in a survey verbal scales for measuring SWB have been used that have not (yet) been included into a corresponding Scale Interval Study, the application of the reference distribution method (Section 6/2.10) should be considered. Otherwise, the WDH-report should be confined to the corresponding statistics of the sample distribution with the explicit NOTE that these statistics concern the *sample distribution* only and are in general not suitable for application in further research. These statistics have an indicative value only.

5. Measurement of happiness and other aspects of SWB is stimulated to be done by using discrete numerical scales. The preferred primary scale is the 11-point numerical scale {0, 1, 2, …, 10}.

6/3.2 Coding of the index numbers in the WDH

The statistics of happiness distributions over the 0 to 10 interval are coded in different ways, depending on how they have been obtained. Since the estimation of the standard deviation is strictly connected with the procedure for the mean, we will focus to the mean completely.

In Section 6/2.1 two mainstreams haven been distinguished to obtain the desired estimates of the population distribution statistics.

In the first one the sampling distribution is declared to act as the population distribution as well. The estimated population mean is computed as the **w**eighted average of the secondary ratings, each with the corresponding observed frequency in the sample as its weight. The approach of this main stream is referred to as the "Weighted Average Approach". All statistics obtained in this way is given "W" as the first letter of their code.

These secondary ratings are always obtained from the primary ones by transformation. This transformation can be based either on the **r**anking of the response options according to increasing ordinal numbers (second letter "R") or on the basis of the judgment of the **s**emantics of the verbal label of the response option (second letter "S").

In the first case a distinction is made between **h**armonized (third letter "h") if linear transformation is applied, and **u**nharmonized (previously labeled as "untransformed"; third code letter "u";). "Harmonized" in this context means that the lowest and the highest secondary are equal to 0 and 10 respectively (Section6/1.4).
If the secondary ratings are obtained on the basis of semantics, this transformation is nonlinear and the result is always harmonized. Again two types can be distinguished, using the "f" and "c" respectively.

The code "WSf" refers to the situation in which the wording of the primary rating is adopted as **f**ixed, irrespective the number, the wording and the position of all other response options in the measure. An example are the Thurstone ratings (Section 6/1.5), where "very happy" is transformed directly into the value 9.3 as its secondary rating under all circumstances.

The code "WSc" is assigned when the semantic judgment is made in the **c**ontext of the total happiness measure, taking into account the position and the wording of all other response alternatives. Such statistics emanate from, e.g., the Happiness Scale Interval Study when the mid-interval values are chosen as the secondary ratings (Section 6/1.6).

The other mainstream is based on the **C**ontinuum Approach and a latent continuous happiness variable, for which a tractable distribution model is assumed (Section 6/2.8). All codes in this category start with the letter "C".
Since in this approach no secondary ratings are defined at all, no transformations are involved nor a weighted average of these can exist.
One of these models is the **s**emi-**c**ontinuous model, where the cumulative frequency distribution is represented with a broken line through all sample observations, as has

been described in Section 6/2.8. The estimated population statistics in this case are given the code “Csc”.

The fully continuous distribution is based on the **b**eta distribution (“B”) and has two variants. In one, the boundaries between the response options is determined either by the opinion of judges for v erbal scales (“CBv”) or based on equidistancy for **n**umerical scales (“CBn”). In the other variant (“CBr”) , the boundaries of the verbal scale have been obtained by calibration according to the **r**eference distribution method as described in Section 6/2.10.

This coding enables one to make a distinction between the estimated population mean when the scale of measurement is the discrete 0 to 10 scale. The code of the traditionally obtained value is WRh; if the continuum approach is followed (Section 6/2.9), the code is CBn.

Schematic survey of estimated population statistic codes:

I. Based *on transformation of the response options*.
Population distribution and sample distribution are identical.
Estimated population mean = Weighted average of secondary ratings.

Transformation:	*secondary ratings*:	*Code*:
Ranking-based (WR·)	harmonized	WRh
	unharmonized	WRu
Based on semantics (WS·)	fixed	WSf
	contextual	WSc

These methods to estimate the population mean and standard deviation are referred to as variants within the *Weighted Average Approach*. The corresponding standard deviation is estimated according to the traditional procedure.

II. Based on the *Continuum Approach* and the cumulative distribution.
Postulated latent variable in population with continuous distribution

Model (C··):	*primary ratings*:	*Code*:
Semi-continuous		Csc
Fully continuous (B-distribution)	numerical-equidistant	CBn
	verbal (judges)	CBv
do. reference method	verbal	CBr

Selected bibliography to Chapter 6.

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: 275-300, DOI: 10.1007/s11205-013-0355-6

Kalmijn, W.M. (2010).
Quantification of Happiness Inequality
PhD-thesis, Erasmus University Rotterdam, The Netherlands: Ipskamp Drukkers, Enschede. Available at http://repub.eur.nl/res/pub/21777/

Kalmijn, W.M. & Arends, L.R. (2010),
Measures of Inequality: Application to Happiness in Nations,
Social Indicators Research **99**(1), 147-161,
ISSN 0303-8300 (print), 1573-0921 (on line).
DOI 10.1007/s11205-010-9573-z

Kalmijn W.M., Arends L.R. (2011)
Erratum to: *Measures of Inequality: Application to Happiness in Nations*
Social Indicators Research, **100**, (2), p.367
ISSN 0303-8300 (print), 1573-0921(on line)
DOI 10.1007/s11205-010-9706-4

Kalmijn, W.M., Arends, L.R. & Veenhoven, R. (2011).
Happiness Scale Interval Study, Methodological Considerations.
Social Indicators Research, 102(3), 497-515,
DOI: 10.1007/s11205-010-9688-2

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

Veenhoven, R. (1993),
Happiness in Nations
Studies in Socio-Cultural Transformation, nr. 2,
RISBO, Erasmus University Rotterdam, (NL).

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, The Netherlands: Springer Press

Appendix A

Estimating the model parameters of the semi-continuous distribution of the latent happiness variable on the basis of numerical scales of measurement.

The model is based on a latent happiness *H* as a real-valued continuous random variable on the closed interval $[0, 10] \subset \mathbb{R}$.
This domain of *H* is partitioned into *k* intervals of equal width =10/*k*.

Within each of these *k* intervals *H* is distributed uniformly, so the density function *g*(*h*) of *H* is

[1] $g(h) = \pi_j/[10/k]=[k/10]\pi_j$ for $(j-1)(10/k)<h \leq j(10/k)$ j=1(1)k
with $\forall j \;\; 0 \leq \pi_j \leq 1$ and

[2] $$\sum_{j=1}^{k} \pi_j = 1.$$

This distribution is called "semi-continuous", because the random variable *H* is continuous, but its density *g*(*h*) is a step function, which is not continuous at the step values $H = j(10/k)$ for j=0(1)k.

The expected value of this variable *H*, denoted μ, is:

[3] $$\mu := \boldsymbol{E}\{H\} := \int_0^{10} hg(h)dh = \sum_{j=1}^{k} \int_{(j-1)(10/k)}^{j(10/k)} \left(\frac{k}{10}\pi_j\right) hdh\,,$$

or

[4] $$\mu = \frac{k}{10}\sum_{j=1}^{k} \pi_j \left(\frac{10}{k}\right)^2 \left(\tfrac{1}{2}j^2 - \tfrac{1}{2}j^2 + j - \tfrac{1}{2}\right) = \frac{10}{k}\left[\sum_{1}^{k} j\pi_j - \tfrac{1}{2}\right]$$
$$\Leftrightarrow \;\; \sum_{1}^{k} j\pi_j = \frac{k}{10}\,\mu + \tfrac{1}{2}\,,$$

where "***E***" is the symbol of the (mathematical) expectation operator.

The variable *H* has a variance:

[5] $$\sigma^2 := \boldsymbol{E}\{(H-\mu)^2\} = \boldsymbol{E}\{H^2\} - \mu^2.$$

$$\boldsymbol{E}\{H^2\} := \int_0^{10} g(h)h^2dh = \sum_{1}^{k} \int_{(j-1)(k/10)}^{j(10/k)} \left(\frac{k}{10}\pi_j\right) h^2dh =$$

$$= \frac{k}{10} \sum_{1}^{k} \pi_j \left(\frac{10}{k}\right)^3 \left(\tfrac{1}{3} j^3 - \tfrac{1}{3} j^3 + j^2 - j + \tfrac{1}{3}\right)$$

$$= \left(\frac{10}{k}\right)^2 \left[\sum_{1}^{k} j^2 \pi_j - \sum_{1}^{k} j \pi_j + \frac{1}{3} \sum_{1}^{k} \pi_j\right]$$

or:

[6] $$\boldsymbol{E}\{H^2\} = \left(\frac{10}{k}\right)^2 \left[\sum_{1}^{k} j^2 \pi_j - \left(\frac{k}{10}\mu + \tfrac{1}{2}\right) + \tfrac{1}{3}\right] = \left(\frac{10}{k}\right)^2 \sum_{j} j^2 \pi_j - \frac{10}{k}\mu - \frac{1}{6} ,$$

From [4], [5] and [6], it follows that:

[7] $$\sigma^2 = \left(\frac{10}{k}\right)^2 \sum_{1}^{k} j^2 \pi_j - \mu^2 - \frac{10}{k}\mu - \tfrac{1}{6} \iff \sum_{1}^{k} j^2 \pi_j = \left(\frac{k}{10}\right)^2 \left[\sigma^2 + \mu^2 + \frac{10}{k}\mu + \tfrac{1}{6}\right]$$

Now we focus on the sample. The *k* counted relative frequencies $\{f_j \mid j = 1(1)k\}$ are to be considered as unbiased estimators $\{\hat{\pi}_j\}$ of the corresponding $\{\pi_j\}$, so:

[8] $$\boldsymbol{E} f_j = \boldsymbol{E} \hat{\pi}_j = \pi_j \qquad j = 1(1)k.$$

In the particular case *k*=10, the sample average happiness value:

[9] $$m := \sum_{1}^{k} j f_j$$

as an estimator of μ is positively biased in the above model in view of [4] and [8]:

[10] $$\boldsymbol{E} m = \sum_{j} j \, \boldsymbol{E} f_j = \sum_{j} j \, \pi_j = \frac{k}{10}\mu + \tfrac{1}{2} = \mu + \frac{1}{2} \iff \boldsymbol{E}\left\{m - \tfrac{1}{2}\right\} = \mu$$

so this bias can be removed easily by replacing *m* with *m*–½ as the estimator.

The sample variance s^2 is calculated as:

[11] $$s^2 = \frac{N}{N-1} \sum_{j=1}^{k} f_j (j - m)^2 ,$$

so

[12] $$\frac{N-1}{N} s^2 := \sum_{j} f_j (j - m)^2 = \sum_{j} j^2 f_j - 2m \sum_{j} j f_j + m^2 \sum_{j} f_j$$

$$= \sum_{j} j^2 f_j - m^2.$$

Taking expectations:

$$[13]\quad \frac{N-1}{N}\boldsymbol{E}s^2 = \sum_j j^2 \boldsymbol{E}f_j - \boldsymbol{E}m^2 = \sum_j j^2\pi_j - \boldsymbol{E}m^2$$
$$= \left(\frac{k}{10}\right)^2 \left[\sigma^2 + \mu^2 + \frac{10}{k}\mu + \tfrac{1}{6}\right] - \boldsymbol{E}m^2.$$

$$[14]\quad \text{var}\,\{m\} = \boldsymbol{E}m^2 - (\boldsymbol{E}m)^2 = \frac{\boldsymbol{E}s^2}{N} \;\Rightarrow\; \boldsymbol{E}m^2 = (\boldsymbol{E}m)^2 + \frac{\boldsymbol{E}s^2}{N} = \left(\frac{k}{10}\mu + \tfrac{1}{2}\right)^2 + \frac{\boldsymbol{E}s^2}{N}.$$

Substitution of [14] in [13] results in

$$[15]\quad \frac{N-1}{N}\boldsymbol{E}s^2 + \frac{1}{N}\boldsymbol{E}s^2 = \left(\frac{k}{10}\right)^2 \left[\sigma^2 + \mu^2 + \frac{10}{k}\mu + \tfrac{1}{6}\right] - \left(\frac{k}{10}\right)^2 \mu^2 - \frac{k}{10}\mu - \tfrac{1}{4},$$

or

$$[16]\quad \boldsymbol{E}s^2 = \left(\frac{k}{10}\right)^2 \sigma^2 - \tfrac{1}{12} \;\Leftrightarrow\; \sigma^2 = \left(\frac{10}{k}\right)^2 \left[\boldsymbol{E}s^2 + \frac{1}{12}\right]$$

In the particular case k=10, s^2 is a negatively biased estimator of the population variance σ^2 in the above model, but this bias can be removed easily by replacing s^2 with $s^2 + {}^1/_{12}$ as an estimator of the population in the semi-continuous distribution model. This bias occurs when ignoring a variance component, to be included in order to account for the within category variability of all k intervals together.

Sometimes, numerical scales are applied with the rating “0” instead if “1” for the most unhappy happiness situation. The most frequently applied scale in this class is the eleven-step scale, which are usually applied as {0, 1, 2 ….., 10} instead of as {1, 2, 3, …,11}. If the cumulative distribution on the basis of the primary distribution is shifted upward over a distance =+1, one gets the distribution on the basis of a {1,2,3,…,11} scale, to which the above formulae apply for k = 11. The only adaptation is that the sample average value m (on the basis of the primary 0-10 scale) is to be replaced with its value m+1 after the shifting.

Application of Eq. [10] to this case gives:

$$[17]\quad \boldsymbol{E}(m+1) = \frac{k}{10}\mu + \tfrac{1}{2} \;\Rightarrow\; \mu = \frac{10}{k}\left(\boldsymbol{E}m + \tfrac{1}{2}\right)$$

So the mean value in the case k = 11 can be estimated without bias by augmenting the average value m on the primary 0 to 10 scale by ½, followed by multiplication by a factor ${}^{10}/_{11}$.

The shift of the distribution does not affect its dispersion, so Eq [16] with k = 11 also applies to the conversion on the basis of the {0, 1, … ,10} primary scale.

The effect of the conversion is demonstrated in the Table 1 below. In the left-hand column, the measured sample statistics m and s on the discrete 0-10 primary scale are given and in the right-hand column the estimates of the corresponding statistics of the latent happiness distribution in the population on a [0, 10] continuum.

Table A1
Discrete and continuous 1-10 scale

	Primary Scale sample Discrete	Estimate population continuous
Mean value	0	0.45
	1	1.36
	2	2.27
	3	3.18
	4	4.09
	5	5.00
	6	5.91
	7	6.82
	8	7.73
	9	8.64
	10	9.55
Standard Deviation	1.5	1.39
	2.0	1.84
	2.5	2.29
	3.0	2.74
	3.5	2.19

Key: If the sample average on a discrete 0-10 scale is 7, then the estimated population mean value on a [0, 10] continuum = 6.82