Happiness in Nations, Introductory text

Chapter 7
HOW THE DATA ARE HOMOGENIZED

Intro All data reported in this book are based on *acceptable* indicators of happiness. As noted in chapter 4, we selected only studies that used queries on happiness that passed a test for face validity. Though all acceptable, the queries used in these studies are not all the same. The queries not only measure different happiness-variants, but also do so by means of different methods. Therefore, the data are not simply comparable.

This heterogeneity has some advantages: The dissimilarity in happiness-variants allows a differentiated look at public happiness. For example, possible inconsistencies between hedonic level and contentment may provide clues about the processes that underlie differences in overall evaluation of life. The diversity in measurement methods is useful as well. It prevents dependency on only one method and provides an empirical basis for estimating method-effects.

Clearly, the heterogeneity has also disadvantages. The main goal of this inventory is to compare public happiness between nations and through time. The more heterogeneous the data, the less we can compare.

Below we will first consider the heterogeneity of the data in more detail (Section 7/1). On that basis we will group the data in subsets of (almost) identical indicators, within which we deem comparison possible (Section 7/2). Next we will consider the possibilities for comparison across these subsets, by means of conversion procedures. Three possible methods are described and tested (Section 7/3).

7/1 HETEROGENEITY IN ACCEPTED QUERIES ON HAPPINESS

As we have seen in chapter 4, happiness has been measured in many different ways. That variety was considerably reduced by the drastic selection for face-validity reported in section 4/2.2. Still we are left with a great amount of subtly differing queries. Exhibit 7/1 presents some illustrative cases.

Exhibit 7/1.
Some current single questions on happiness

Question	*Our classification*
Taking all together, how happy would you say you are..? Very happy, pretty happy, not too happy	O-HL
Generally speaking, how satisfied are you with the life you lead…? Very satisfied, fairly satisfied, not so satisfied, dissatisfied	O- SLL
Here is a picture of a ladder. Suppose that the top represents the best possible life for you, and the bottom the worst possible life. Where on the ladder do you feel you stand at the present time?	O-BW
How do you feel about your life as a whole…? Delighted, pleased, mostly satisfying, mixed, mostly dissatisfying, unhappy, terrible?	O-DT
Are you most of the time..? In very good spirits, in good spirits, in low spirits, in very low spirits	A-AOE
How would you rate yourself as to how successful you have been in terms of achieving your goals and aims in life…? Think of 10 s very successful and 0 as being entirely unsuccessful	C-RA

In the earlier discussion on measurement of happiness (section 4/3) we have seen that accepted queries can differ in the following respects.

- Focus: The queries address different kinds of happiness.
- Time reference: Questions about happiness refer to different periods
- Mode of assessment: Various questioning techniques have been used, and hedonic level has also been assessed by observation
- Rating of the response: Responses have been recorded on different kinds of rating scale of different length. There are also differences in the wording of verbal labels.

Below we will consider to what extend these variations in measurement jeopardize comparison across nations and trough time.

7/1.1 Heterogeneity in focus

Though all queries in exhibit 7/1 concern happiness, they do not tap quite the same kind. Hence the data they yield are not quite comparable: we cannot say that inhabitants of country A are happier than inhabitants of country B, if we know only that the former score high on an indicator of contentment and the latter low on an indicator of hedonic level. As explained in chapter 4, these variants do not necessarily coincide: resignation may involve high contentment together with depressed mood. Therefore this database does not throw all happiness on one heap, but presents the findings by happiness variants.

Within main variants of happiness there are further limitations to comparability. Questions on the same kind are phrased differently and these differences are sometimes too great to allow meaningful comparison. For example: in the same population, the average answer to a question on how 'happy' one is (type O-HL) can be more positive than to a question on how close life is to 'best possible' (type O-BW). That means that the former interrogation method gives a more favorable estimate of true happiness in that population than the latter. Likewise, variations in focus can complicate comparison between scores on indicators of hedonic level. It is for instance not at all sure that hedonic level is higher in a country were a single question on general mood (type A-OE) is rated 7.5 on a 0-10 scale, than in a country were average Affect Balance (A-BB) is 6.5 on the same scale or, interviewer rated cheerfulness (A-CA) is 5.5.

Exhibit 7/1.2
Variation in time reference in questions on happiness-in-life (type O-HL)

How happy would you say you have been up to now?

Taken all together, how happy would you say you are?

How happy do you feel as you live now?

Generally speaking, are you a happy person?

How happy is your life at this moment?

7/1.2 Heterogeneity in time reference

One of the most commonly used indicator is a single closed question on 'happiness-in-life' (type O-HL). This question has been put in different ways, which differ in time-reference. Exhibit 7/1.2 presents some examples.

Though all these questions use the word 'happiness' as the key-term, they differ subtly in time-perspective ('up to now', 'generally','now', 'at the moment'). Such minor differences can possibly produce small variations in average scores, which could jeopardize comparisons between nations and through time. A check of that possibility is planned. For the time being it is assumed that variations in time reference do not seriously affect the scores.

7/1.3 Heterogeneity in assessment modes

Next to this substantive variation there are difference in methodology. One of these differences is in the mode of assessment. The most common technique is interrogation, but hedonic level has also been assessed by cheerful appearance (A-CA). These techniques are to different to allow meaningful comparison. A self estimate of 6 on a 10 step scale could correspond with an external rating of 4. The relation between these appraisal methods is not sufficiently investigated as yet.

Interrogation is typically done by means of standard questions, with precoded response options. Mostly these are single questions (code sq), sometimes asked twice in the interview and the ratings are added (coded sqt). Such single questions are deemed comparable, provided that led item and response options are equivalent.

Next to single questions there are also multiple item questionnaires. Average scores on such inventories cannot be meaningfully compared with average responses on a single question, even when both are expressed on a same numerical range. For instance, when in a country the average score on a ten step item on general mood (A-AOE) is 6, the mean score on the 10 item Bradburn's Affect Balance Scale (A-BB) could be 7.

7/1.4 Heterogeneity in rating of responses

All queries rate the degree of happiness on some scale. These rating scales differ also in some respects. In the rating device, in the number of response categories and in the precise labeling of response options.

Kind of rating scale

Most queries use verbal rating scales, such as 'very happy', 'fairly happy', 'not very happy' and 'unhappy'. Average scores are computed by attributing numerical values to these options, f.e. by giving 'very happy' responses the value '4' and 'unhappy' value '1'. Another technique is to have the respondents themselves express their happiness in a number, mostly between 1 and 10. Clearly, such values are not comparable, not even when expressed on the same range.

Number of response categories

Questions type O-HL are commonly presented with three answer categories. However, there are also variants with four or five answer categories. Exhibit 7/1.4a presents some examples of rating-scales of varying length. Clearly such differences hamper comparison

as well. One cannot say that public happiness is higher in a country with score 2 on a 3-step scale, than in a country with score 3 on a 5-step scale.

Exhibit 7/1.4a
Variation in number of response categories with a simple closed question on happiness

---

Taken all together, how happy would you say you are

very happy	very happy	very happy
fairly happy	quite happy	rather happy
not too happy	not very happy	neither happy nor unhappy
	not at all happy	fairly unhappy
		very unhappy

Labeling of verbal response categories
The common lead question 'In general, how happy would you say you are' is typically presented with three answer categories. However, these answer categories are not always identically labeled. Exhibit 7/1.4b presents some examples.

The difference is in this case largely in the last answer category: 'not at all happy', 'unhappy', 'not very happy' and 'not too happy'. If all these response options are given an equal weight (i.e. 1 on a 1-3 scale), questions that provide a response option of the former kind will probably yield a more favorable estimate of true happiness in a population than the latter. The more pertinent the unhappy category, the less respondents will choose it, and the more will characterize themselves as 'fairly' happy. Similar

Exhibit 7/1.4b
Variation in labeling of response categories with a 3-step question on happiness-in-life (type O-HL/g/v/3)

In general, how happy would you say you are?

very happy	very happy	very happy
fairly happy	fairly happy	pretty happy
not at all happy	not happy	not too happy

---

variation in response categories exists with other question types, in particular with questions on life-satisfaction (type O-SL). Obviously, such variations also limit the comparability of the data, especially if items also diverge in the number of response categories offered.

7/2 GROUPING COMPARABLE FINDINGS

The main goal of this database to compare public happiness between nations and through time. Therefore, its data are organized in comparable sets. These sets are the tables in this collection. The construction of these comparable sets involved the following steps:

7/2.1 Grouping by focus of query

As we have seen in section 7/1.1, there are differences in focus at two levels: firstly between the in happiness variant addressed and secondly in the phrasing of lead questions. Since these differences impede comparison, the findings yielded with such queries are presented separately.

Main variant of happiness

differences make do not produce comparable results. In the same country at a particular time, overall happiness may rate point 7.5 on a 0-10 scale, hedonic level 6.0 and contentment 8.2. For that reason this database presents the data of different happiness variants separately. It has four main sections: section 1 presents the findings on 'overall happiness', section 2 findings on 'hedonic level' and section 3 findings on 'contentment'. Because classification is dubious in some cases, section 4 presents findings yielded with 'mixed indicators' separately. This grouping breaks the collection into four pieces: one big piece (overall happiness, code O), one smaller piece (affect level, code A) and two minor pieces (contentment (C) and mixed queries (M).

Minor variation in key words

In section 7/1.1 we have also seen that queries on the same happiness variant can differ slightly. In the case of overall happiness the difference is in the key terms such as 'happiness', 'life-satisfaction', 'best-worst' or 'delighted-terrible'. Comparison between scores on such different items is not possible either. Therefore, the data are split up further by focus variant. This breaks the collection into nine pieces: four variants of overall happiness (O-HL, O-SL, O-BW, O-DT), three variants of hedonic level (A-AOE, A-BB and A-CA), one variant of contentment (C-RA) and one mixed indicator (M-AO). Within each main focus category, there is also a table of miscellaneous items, which brings the total on twelve tables.

7/2.2 Grouping of near-identical items

Still there are differences between queries that focus on the same subject matter, in

particular among single questions about overall happiness. The difference is now in the response scales.

Happiness-in-life Single questions on happiness-in-life (O-HL) were further differentiated in three more homogenous classes and a rest category. The three class distinction is mainly based on the number of response categories, but we also considered similarity of lead phrase and category labels. Divergent items were moved to the rest category. This resulted in four tables: Table 1.1.1a: '3-step happiness', table 1.1.1b: '4-step happiness', table 1.1.1c: '5-step happiness' and table 1.1.2: 'further single questions on happiness'; the rest category.

Though almost identical, the items in the first three homogenous categories still differ slightly in lead question and labeling of answer categories. To allow further differentiation, the tables contain references as to the precise wording of the questions. For instance, table 1.1.1a enumerates twelve variations on the question 'In general, how happy would you say you are?' These variations are marked by a code: HL1 to HL12. In the notes to the table, all questions are presented in English translation.

Satisfaction-with-life In the case of single questions on satisfaction-with-life (O-SL) there are two main question types: Firstly, questions that focus on 'satisfaction with the life one leads' and that use short verbally labeled answer categories (code O-SLL). Secondly, questions about 'satisfaction with life-as-a-whole' that are represented on a longer graphic rating scale of which only the extremes are defined verbally (O-SLW). Within these two variants there are further differences in length of rating-scales.

Again, identical subsets were created on the basis of similarity in lead question and rating-scale. This resulted in five separate tables. First, three tables on the question on 'satisfaction with the life one leads'; table 1.2.1a: 3-step way-of-life-satisfaction, table 1.2.1b: 4-step way-of-life satisfaction and table 1.2.1c: 5-step way-of-life satisfaction. Next, two tables on 'satisfaction with life-as-a-whole': table 1.2.2a: 10-step life-satisfaction and table 1.2.2b: 11-step life-satisfaction. Items that do not fit any of these subsets are again separately presented in a rest category (table 1.2.3 various life-satisfaction items).

The resulting classification of the data is presented in exhibit 7/2.3. That exhibit also mentions the tables in this database where the scores on these queries are reported. The table titles in the exhibit are printed in italics: ***bold italics*** refer to similar items; here comparison is possible between nations and through time. *Non-bold italics* denote heterogeneous rest-categories, that do not allow comparison.

Exhibit 7/2.3 does not enumerate the queries that have been used in cross-national studies on specific groups, such as university students.

7/2.3 Equivalent items

In the foregoing section we have grouped these items in near-identical classes. Accordingly, the scores on these indicators are presented in separate table and hence marked as incomparable. In some cases that is too strict however. There are clusters of items that involve essentially the same question, and differ only slightly in number and labeling of response categories. Though the numerical scores on these questions are not comparable, their content is equivalent. Therefore they are suitable for conversion to a

same standard. We call this *equivalent items*. Conversion procedures will be discussed in section 7/3.3.

Exhibit 7/2.3
Classification of indicators of happiness used in representative nation surveys according to comparability

Code	short name		table
OVERALL HAPPINESS			
O-HL	Single closed questions on happiness-in-life		
	3-step happiness	}	***1.1.1a***
	4-step happiness	} ***equivalent***	***1.1.1b***
	5-step happiness	}	***1.1.1c***
	Further question on happiness		
	Various happiness items		*1.1.2*
O-SL	Single closed questions on satisfaction-with-life		
	• ***3-step satisfaction***	}	***1.2.1a***
	• ***4-step satisfaction***	} ***equivalent***	***1.2.1b***
	• ***5-step satisfaction***	}	***1.2.1c***
	• ***10-step satisfaction***	}	***1.2.2a***
	• ***11-step satisfaction***	} ***equivalent***	***1.2.2b***
	Further questions on satisfaction-with-life		
	• *various life-satisfaction items*		*1.2.3*
O-BW	Single closed questions on Best-Worst possible life		
	• ***11-step Best-Worst (Cantril ladder)***		***1.3***
O-DT	Single closed questions on Delighted-Terrible life		
	• ***7-step Delighted-Terrible***		***1.4***
HEDONIC LEVEL OF AFFECT			
A-AOE	Single closed questions on usual affect		
	• *Various feeling items*		*2.1*
A-BB	Multiple closed questions on specific affects in the recent past		
	• ***10-item Affect Balance Scale***		***2.2***
A-CA	Rating of cheerful appearance		
	• ***8-aspect rating of cheerfulness***		***2.3a***
	• ***global rating of cheerfulness***		***2.3b***
CONTENTMENT			
C-RA	Single closed question on perceived realization of wants		
	• *Various contentment items*		*3*
MIXED INDICATORS			
M-AO	Questions that concern both overall happiness and hedonic level		
	• Various items		*4*

Bold italics : comparable items *Non-bold italics* : non-comparable items
Equivalent : comparisons across tables possible after transformation

Most of the questions on happiness-in-life (O-HL) are considered equivalent. Among the questions about satisfaction-with-life two groups of equivalent items are discerned: Firstly, questions about satisfactions with ones way of life (O-SLL), which are typically scored on short rating scales. Secondly, questions about satisfaction with life-as-a-whole that are rated on longer graphic scales (O-SLW). See again exhibit 7/2.3.

7/3 CONVERTING SCORES ON NON-IDENTICAL ITEMS

This partition of the data into sets of (near) identical questions breaks this data collection into splinters. The number of cases for comparisons is thereby reduced considerably. Though comparison is better possible within the purged categories (tables), there is less to compare: less countries in each subset and less years in time-series. Therefore, we considered the possibilities for converting scores on different indicators to a common standard. We focussed on transforming averages. Conversion of measures of dispersion was not attempted.

Converting average scores on different questions on happiness is in fact estimating how respondents would have answered on a question that was not presented to them. That estimate is made on the basis of their responses to one or more other questions they did answer.

The simplest estimate of that kind is inferring the average response on an (unasked) question A in the light of a response on question B. For instance: we can assume that people in a country which scores 5 on a 10-step life-satisfaction question (B) would have scored 2.5 on a 5-step happiness (A) item. Such estimates can help to reduce the number of 'missing values' in nation-sets or time-series. If we have too few observation on the basis of question A, we can supplement these with transformed scores on question B.

More far-reaching is transforming all scores for all questions to one standard; i.e. to an imaginary 100 step happiness scale. That would of course create the greatest possible data-set.

Such estimates are no more than guesses. One can never be sure how people in a country would have answered a question that was not posed to them. Still, one can make educated guesses. Lets us take a look at the possibilities for transformation and see how these work out in this data-set.

7/3.1 Converting average scores on measures of different happiness variants

As noted above, the indicators of overall happiness, hedonic level and contentment measure essentially different things. Hence scores on these indicators can *not* be transformed to one common standard. At best such scores can be combined in an overall index. However, that is hardly helpful conceptually. Indicators of overall happiness are already supposed to cover the whole. Moreover, such a procedure would not create more comparable data.

7/3.2 Converting average scores on different measures of a same happiness variant

The four blocks in exhibit 7/1.1 present different methods for measuring the same happiness variants. All indicators type O in the first block are supposed to measure 'overall happiness', all indicators type A in second block 'hedonic level' and all indicators type C in the third block 'contentment'. In principle different measures of the same phenomenon are comparable. However in practice they are not. As we have seen in section 7/1.2, method-effects may veil the differences in true happiness.

Still, it is possible that there is constancy in these method effects which may allow estimates of missing values. Suppose that we have average scores on two questions in a sizable number of nations; i.e. on the question on 'happiness-in-life' (X) and on the question about 'best-worst possible life' (Y). Suppose further that the average scores on the former question are typically more positive than on the latter, and that the relation can fairly well be described by an equation; for instance the formula X = 1,25 + 0,5 Y, where X is 5-step happiness (independent variable) and Y is 11-step best-worst (dependent variable). We can then estimate average best-worst rating of a country of which we have only information about average happiness-in-life. The parameters for such a formula can be found by means of regression analysis. In that way we can derive estimates for missing happiness values in nation-sets and time-series.

Such conversions of one measure to another are risky, because one is never sure that the equation derived from a set of countries for which scores on both measures are available, also applies to the country with a missing X-score. Yet, the risk that the equation does not apply is clearly smaller if it is based on many cases (countries, years) and if the variability (deviance from the regression line) is low.

We explored this possibility in an analysis of cross-national studies that involved several items on overall happiness. We inspected the relation of responses in four pairs of questions: 1) happiness-in-life vs. satisfaction-with-life, 2) happiness-in-life vs. best-worst-possible-life, 3) satisfaction-with-life vs. best-worst-possible-life and 4) happiness-in-life vs. delighted-terrible-life.

The scores on these pairs of items were analyzed by means of bi-variate regression. If there is a clear linear relation, the average scores must neatly fit the regression line. In that case the equation of that line provides a formula for converting one score to another. If however the scores appear to be scattered, there is apparently no consistency in the ratio of responses to these questions across countries.

This analysis requires two choices: a choice for the most appropriate regression line and a choice of the acceptable divergence from that line. The first choice concerns three possible regression lines: 1) the regression line with happiness-Y as the dependant variable, 2) the regression line with happiness-X as the dependant variable and 3) the intermediate line based on the z-scores of Y and X. If we want to predict a missing score on happiness-Y in a particular country on the basis of happiness-X in that country, regression line 1 is clearly more appropriate than regression line 2; because Y is the dependent variable in this case and not X. Regression line 1 is also preferable to z-score line 3.

The acceptable dispersion around the regression line is usually indicated by 5% confidence intervals at each side. These confidence intervals are typically more narrow

around the average than at the extremes of the distribution. Due to the limited number of observations at hand here, it is not well possible to estimate such confidence intervals. Therefore, we reverted to a more simple criterion and fixed the acceptable dispersion at 10% of the possible scale range; 5% above and 5% under the line. This may seem a rather narrow tolerance area, but the actual range of variation in the data at hand here is in fact only 50% of the possible range.

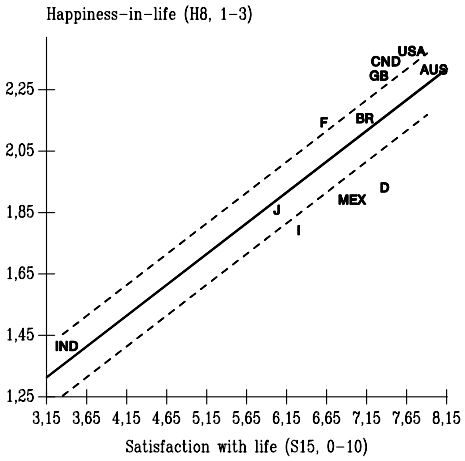
The 10% tolerance area is depicted graphically in the exhibits below. If a sizable amount of the cases are outside that area, the dispersion is clearly too great and consequently transformation not justified. If there are only a few cases slightly outside the area, it is worth considering the probability that these are incidental outliers.

Regression of responses to items on 'happiness-in-life' and 'satisfaction-with-life'

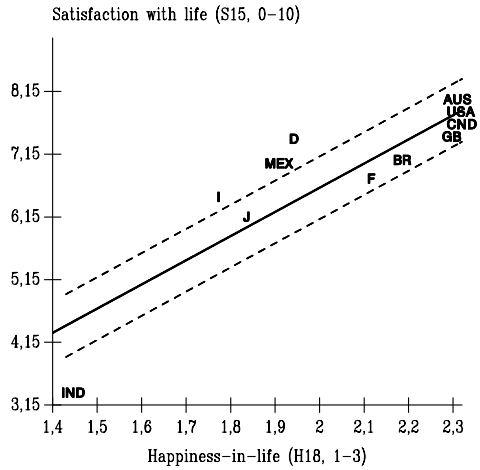
Two cross-national studies involved questions on both 'happiness-in-life' and 'satisfaction-with-life'. The 11 nations Gallup/Kettering World Survey and the 22 nations World Value Study I.

The Gallup/Kettering World Survey involves a 3-step happiness (table 1.1.1a, question O-HL8) and an 11-step life-satisfaction (table 1.2.2b, question O-SLW17). The average scores on these items in the eleven countries were crossed. See the scattergram in exhibit 7.3.2a. Six of the eleven scores are very close to the regression line. Four cases are outside the interval however. The regression line is heavily influenced by the extreme case of India. However, there are no reasons to consider that case as invalid. Moreover, omitting India does not provide a better fit.

Exhibit 7/3.2a
Average scores on items about 'happiness-in-life' and 'satisfaction-with-life' in 11 countries
11 nations in 1975



Happiness = 0.68 + 0.20 Satisfaction.



Satisfaction = -1.10 + 3.86 Happiness.

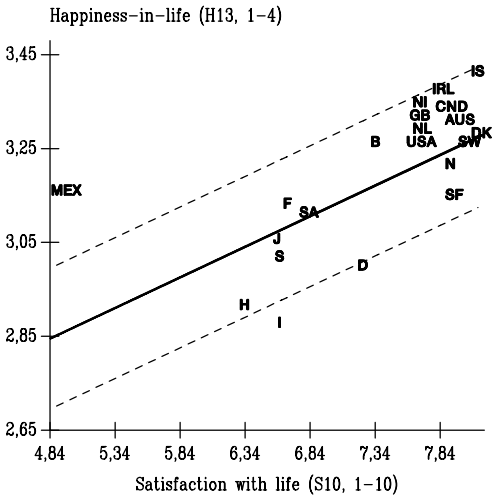
Data: Gallup/Kettering World Survey 1975. See table 1.1.1a and 1.2.2b.
Nation codes: see p 283. r = +.88

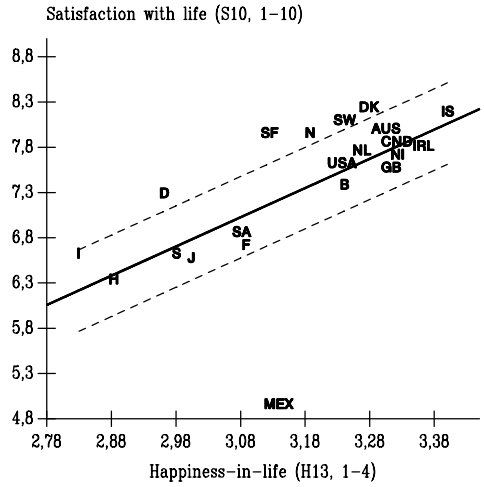
The World Value Survey I involves a 4-step happiness item (table 1.1.1b, question O-

HL13) and 10-step life-satisfaction (table 1.2.2a, question O-SLW10). The average scores on these items are crossed. See exhibits 7/3.2b. Again there are quite a few cases outside the acceptable range of variation.

All in all, these data do *not* provide a solid basis for estimating life-satisfaction in countries on the basis of responses to questions about happiness-in-life, or vice versa. This is a pity, because the present data-set involves many missing values that might have been estimated in this way.

Exhibit 7/3.2b
Average scores on items about 'happiness-in-life' and 'satisfaction-with-life'
22 nations in 1980





Happiness = 2.21 + 0.13 Satisfaction

Satisfaction = -2.92 + 3.23 Happiness

Data: World Value Study I 1980. See table 1.1.1b and 1.2.2a.
Nation codes: See p 283. r = +.65

Regression of responses to questions on 'happiness' and the 'best-worst possible life'

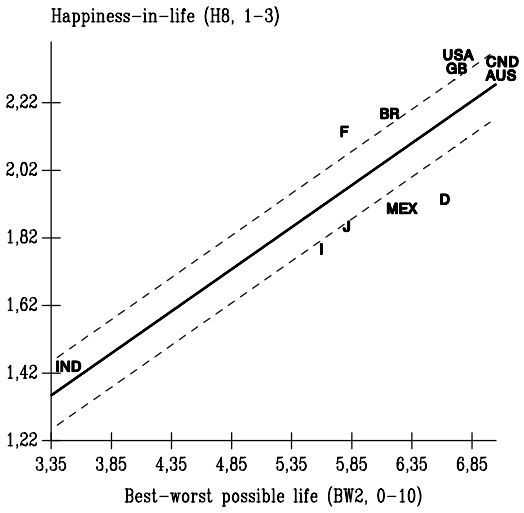
The 11 nation Gallup/Kettering World Survey also involved a question on 'Best-Worst' possible life (table 1.3, question O-BW 2, currently known as the 'Cantril Ladder- rating of present life'). The average scores on this item were also crossed with happiness (table 1.1.1a, question O-HL8). See exhibits 7/3.2c. In this case several cases are outside the interval, though not very far. Again Mexico and West Germany are most deviant. Estimating BW scores on the basis of responses to questions about happiness-in-life is this very risky.

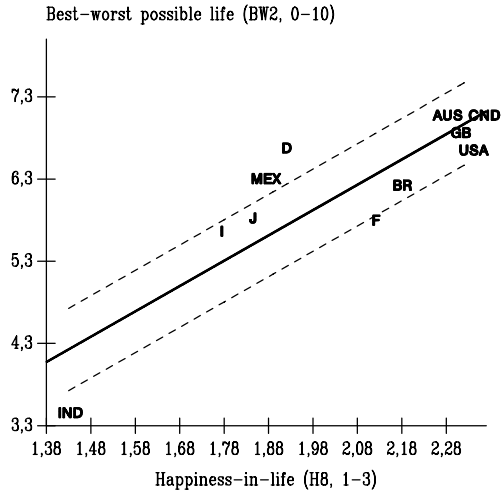
Regression of scores on questions about 'satisfaction-with-life' and 'best-worst possible life'

The above mentioned questions on 'satisfaction-with-life' and 'best-worst possible life' in the Gallup/Kettering World Survey are also crossed. This pair is especially interesting because both questions use the same rating scale: Cantril's 0-10 step ladder-picture. The data are presented in exhibit 7/3.2d. In this case all the scores are neatly within the 10% interval. Unlike the previous scattergram, India and West Germany do not appear as

deviant. This suggests that we can obtain reasonably good estimates of missing Best/Worst scores on the basis of observed Life-Satisfaction scores. The appropriate transformation formulae are mentioned at the bottom of the exhibit.

Exhibit 7/3.2c
Average scores on items about 'happiness-in-life' and 'best-worst possible life'
11 nations in 1975



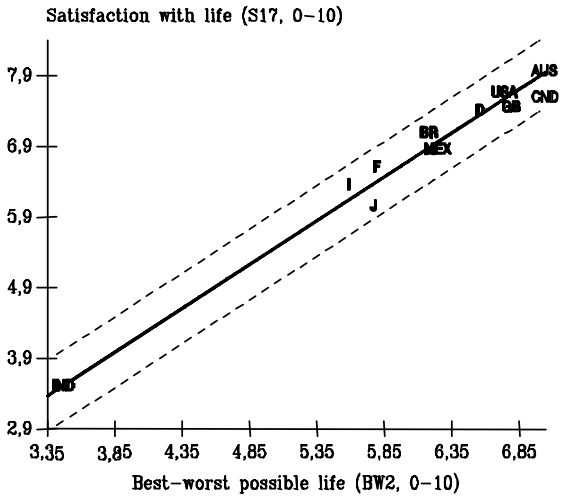


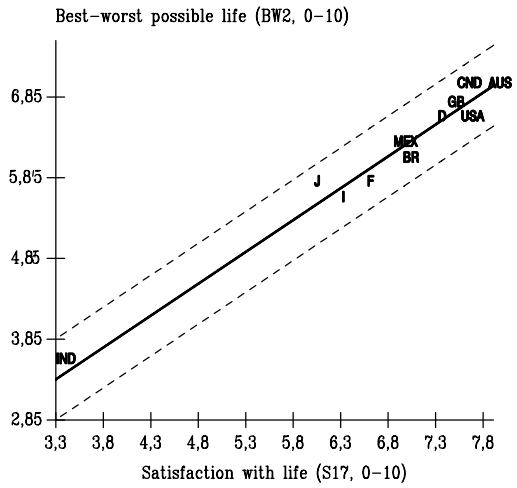
Happiness = 0.52 + 0.25 Best/Worst.

Best/Worst = 0.18 + 3.08 Happiness.

Data: Gallup/Kettering World Survey 1975. See table 1.1.1a and 1.3.
Nation codes: See p 283. r = +.88

Exhibit 7/3.2d
Average scores on questions about 'Satisfaction-with-life' and 'Best-Worst possible life' (both questions rated on same graphic scale)
11 nations in 1975





Satisfaction = -0.79 + 1.24 Best/Worst.

Best/Worst = 0.75 + 0.79 Satisfaction.

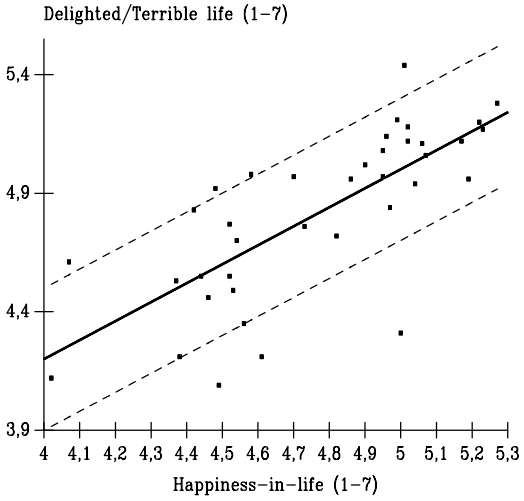
Data: Gallup/Kettering World Survey 1975. See table 1.2.2b and 1.3.
Nation codes: See p 283. r = +.99

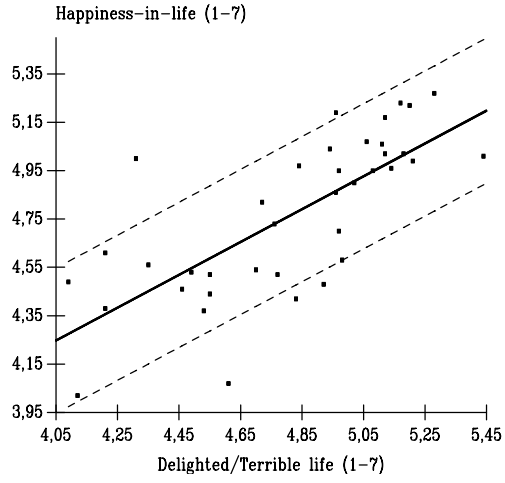
In the present data-set there are many missing values which can be substituted in this way. For a lot of countries we know the score on 11-step satisfaction-with-life around 1975, but not the score on 11-step best-worst possible life: f.e. Austria, Belgium, Finland and the Netherlands. In the latter case conversion yields an estimated Best-Worst score for the Netherlands of 6.8, which is close to the score 7.1 observed in a small sample at that time. There are also quite some countries of which we know the best-worst score at a particular time, but not average satisfaction-with-life. Such cases are Israel, Poland and Yugoslavia. Compare table 1.2.2b with table 1.3.

Regression of responses to question on 'happiness-in-life' and 'Delighted-Terrible-life'
Michalos' 'Global Student Well-Being Survey' involves both a question on 'Happiness-in-life' and a question on feelings about life in terms of 'Delighted-Terrible'. This data set is particularly suited for the purpose of identifying a possible stable ratio in the responses to these items. Firstly, both questions are rated on a 7-step scale. Secondly, the number of nations is largest (38). Thirdly, the respondents are university students and probably understand differences in wording better than respondents drawn from the general population.

The data are presented in exhibit 7/3.2e. Again we see a clear pattern of the scores around a linear regression line, but once more there are outliers. Of the 38 countries, 6 are outside the tolerance interval; two countries are beyond the line in both analyses (Thailand and Bangladesh). Though not dramatic, this deviance marks that transformation of one score to an other is risky.

Exhibit 7/3.2e
Average scores on 'happiness-in-life' and 'delighted-terrible life' (Both questions rated on 1-7 scales). University students in 38 countries 1985





D/T = 1.00+ 0.80 Happiness.

Happiness = 1.50 + 0.68 D/T

Data: Michalos (1986). Student Well-being Survey. r = +.74

In the present data-set there is only one case of a missing value, which could be estimated in this way. That is the case of Russian university students, of which we have a D/T score (4.15 on a 1-7 scale, but not a happiness rating. The estimated happiness score would be 4.32.

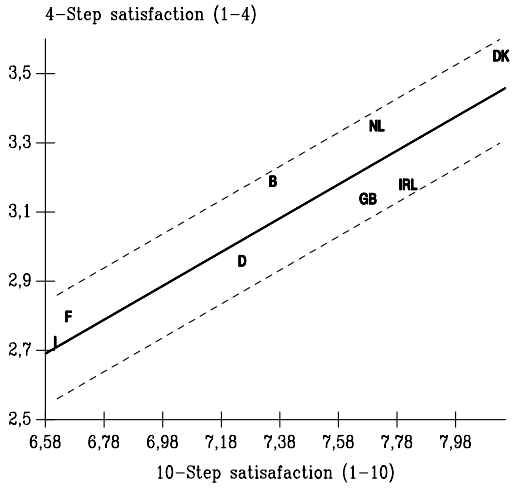
Regression of responses to 4-step 'satisfaction with way-of-life' and 10-step 'satisfaction with life-as-a-whole'

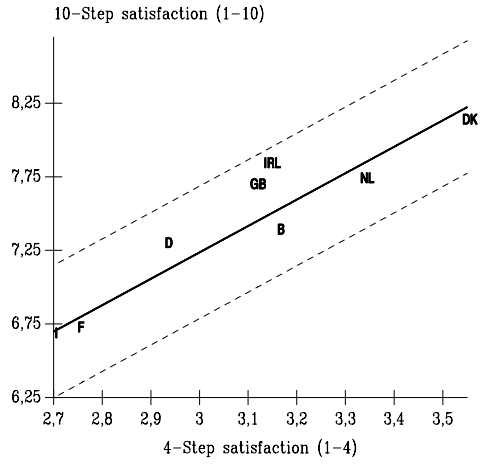
Finally we considered two items on 'satisfaction-with-life'. Though there is no study that asked both questions in the same interview, there are highly comparable data from surveys in West-European countries in 1981. The Eurobarometer survey involves a 4-step question on satisfaction with the life one leads (Question O-SLL4, table 1.2.1b). World Value Study I contains a 10-step question on satisfaction with life-as-a-whole (QuestionO-SLW12, table 1.2.2a). Eight countries were involved in both studies.

The scores on these items in the eight countries are presented in exhibits 7/3.2f. The correlation is high (r=+.94) and the scores are all within the 10% interval. Hence transformation seems justified in this case. Missing values on 4-step satisfaction with way of life can be estimated on the basis of scores on 10-step satisfaction with life-as-a-whole. Reversibly, missing values on 10-step satisfaction with life-as-a-whole can be estimated reliably on the basis of observing responses to 4 step satisfaction with way-of-life. The conversion formulae are again presented at the bottom of the exhibits.

Several missing values can be substituted in this way. Of Greece and Luxembourg we have scores on 4-step satisfaction in 1981 (table 1.2.1b) but not scores on 10-step satisfaction (table 1.2.2a). Conversion of 4-step satisfaction yields the following estimates of 10-step satisfaction: Greece 6.44 and Luxembourg 7.84. Likewise we can now estimate 4-step satisfaction of many countries of which we have only 10-step satisfaction scores, such as Australia, Hungary, Iceland, Mexico and White Russia.

Exhibit 7/3.2f
Average scores on 4-step 'satisfaction with way-of-life' and 10-step 'satisfaction with life-as-a-whole'
8 nations in 1980/81





4-step satisfaction = - 0,53 + 0,49 10-step

10-step satisfaction = 1,85 + 1,8 4-step

Data: Eurobarometer 1981. See table 1.2.1b. World Value Study I. See table 1.2.2a.
Nation codes see p 283. r = +.94

In summary

In two cases (pairs of items) the consistency of responses is sufficiently great. These pairs are: 1) 11-step 'satisfaction-with-life-as-a-whole' and 11-step 'best-worst-possible-life', and 2) 4-step 'satisfaction with way-of-life' and 10-step 'satisfaction with life-as-a-whole'. In these cases transformation to and from seems justified, at least when the transformed values are between the highest and lowest observed score of untransformed happiness (intrapolation). Going beyond the observed range (extrapolation) is not advised.

In three cases the consistency is not sufficient however. The cases concerned are all pairs with 'happiness-in-life': a) with 'satisfaction with life-as-a-whole', b) with 'Best-Worst possible life' and c) 'with Delighted-Terrible life'. Transformation is therefore not recommendable in these cases.

Substitution of missing cases by means of transformation results in identical data-sets for different happiness items. In other words: the concerned tables in part II of this book will contain the same cases (nations-years). This leaves the user a choice: he can use either one table or the other. This option may tempt to go for the one that produces the most desirable results. That is likely to create confusion. We therefore advise to choose the data-set (table) with the least transformed cases. Original scores are always preferable to estimated ones. If both sets might include about equally much of such cases we advise to consider them both, in order to check possible differences.

7/3.3 Conversion of average scores on equivalent items

A more modest approach is to focus on measures of the same kind: that is on similar questions about the same happiness variant. In section 7/2.4 we have already identified items that are 'equivalent' in content, but differ in rating scales and are therefore not comparable. For example: there is no substantial difference in the 3-step and the 4-step variant of the question 'Taking all together, how happy would you say you are' (O-HL7: very/fairly/not-too, O-HL13 very/quite/not-very/not-at-all). Still the numerical scores are not comparable: we cannot say that 2.5 on the former item marks higher happiness than 3.0 on the latter. Conversion is easier in this case. We need not go into comparison of qualities (characterization of life), but can restrict to estimates of quantity (ratings of the same).

As in the previous case, we could try to transform scores by means of regression equations. However, that approach requires that we estimate a linear relationship and establish whether the observations are sufficiently close to the regression line. Unfortunately, we have insufficient data for that purpose (7/3.3.1). Yet, there are more possibilities in this case. As we deal with differences in measures of quantity only, we can try to transform these to a common scale; that is, 'standardize' the average scores. Two methods can be used for that purpose: weighting of response categories by experts and simple linear transformation. Expert-weighting is most appropriate where we want to standardize scores on rating-scales that differ in verbal labels of response categories. This method will be described in section 7/3.3.2. Linear transformation is more appropriate

where the difference is only in the length of graphic rating-scales. That method will be considered in section 7/3.3.3. Next the sections 7/3.3.4 and 7/3.3.5 will check the validity of these transformations.

7/3.3.1 Regression of average scores on equivalent items

In principle we can follow the same method as used for transforming scores from different methods, now hoping for a greater consistency. That procedure requires studies that involve several such subtly differing questions; preferably many of that kind in different nations.

Such studies are hardly available however. The surveys that involve different questions on happiness typically pose questions of different kinds and not variations on the same. Only studies that focus on measurement issues sometimes consider different variations of equivalent questions. Unfortunately, these studies do not cover all the variants at hand here. Even the detailed study on happiness questions by Andrews and Withey (1976) involved only a fraction.

7/3.3.2 Standardization by expert weighting

Quite another approach is to read equivalent items carefully and estimate off hand the level of happiness indicated by the various answer-categories on some common scale. For instance one could consider the common three step happiness item 'Taking all together, how happy would you say you are: very happy, pretty happy or not too happy'. One can then estimate the weight of these three responses on a 0 to 10 step scale. For example: a weight of 9 for the 'very happy' response, 7 for 'pretty happy' and 4 for 'not too happy'.

The weights awarded depend of course on personal interpretation of the questions and on response tendencies. This bias can be reduced by using more than one judge. The more judges, the greater the chance that personal interpretation-differences neutralize each other. The use of more than one judge also makes variation in interpretation visible. Inter judge-reliabilities can be computed and if these are not satisfactory, the attempt can be stopped.

Judges can be 'typical respondents' or 'experts'. The use of respondents has the advantage that one gets a better view on the interpretation of the item in practice. However, that advantage applies only when there is a typical respondent, which is not the case in this comparative endeavor. Ratings can also be made by people who are well acquainted with the matter; for instance students, experienced interviewers or colleague investigators. That latter method is described by Torgerson (1958:67). We followed his directions.

This weighting method is obviously a rather uncertain one. There is no check whether one assigns the right weights or not.

Weighting responses to equivalent questions on happiness-in-life

Nine subtly different questions on happiness-in-life (type O-HL) were considered. The introductory sentence of these questions is almost identical. The difference is in the length and labeling of the rating scales. These items were rated by ten investigators working on the World Database of Happiness, who were all well acquainted with the subject. These experts graded the degree of happiness indicated by each response category on a 0-10 scale. The results are presented in exhibit 7/3.3a.

These ratings appear fairly consistent. Standard deviations are typically below one interval on this ten step scale. Only in the weighting of unhappy categories do the judges diverge. Not surprisingly this occurs on the item that provides only one possibility for expressing unhappiness.

The overall means were used to compute standardized 0-10 scores for all items on happiness-in-life reported in the tables 1.1.1a, 1.1.1b and 1.1.1c. These converted means are reported in a column in the tables next to the original means.

Weighting responses to equivalent questions about satisfaction-with-life

In the same vein response categories of questions on satisfaction-with-life (type O-SL) were weighted. The results are presented in the exhibits 7/3.3c and 7/3.3d. In only two cases can we see how a same answer category is answered in the context of a slightly different configuration of further response options. The difference is negligible in these cases. Hence we decided again to use the overall mean.

The overall means were used to compute a 0-10 range standardized score for all the findings in the tables 1.2.1a, 1.2.1b and 1.2.1c. These standard means are reported in a column in the tables next to the original mean.

Exhibit 7/3.3a
Ratings on a 0-10 scale of response-categories to equivalent questions on happiness-in-life.

Question	*Mean rating*	*Standard deviation*	*table*	*query type*
In general, how happy would you say you are?				
- very happy	9.2	0.75	1.1.1a	H1/2
- fairly happy	6.4	0.49		
- not very happy	3.5	1.28		
In general, how happy would you say you are?				
- very happy	9.2	0.98	1.1.1a	H3
- fairly happy	6.3	0.64		
- not at all happy	1.1	0.94		
In general, how happy would you say you are?				
- very happy	9.3	0.64	1.1.1a	H4/5
- fairly happy	6.5	0.50		
- not very happy/not at all happy*	2.5	1.53		
In general, how happy would you say you are?				
- very happy	9.2	0.98	1.1.1a	H6
- fairly happy	6.3	0.64		
- not happy	2.7	1.27		
In general, how happy would you say you are?				
- very happy	9.0	0.63	1.1.1a	H7/8/9
- fairly happy	6.8	0.60		
- not too happy	4.2	0.75		
Taking all things together, how would you say things are these days				
- very happy	9.3	0.64	1.1.1a	H10/11/12
- pretty happy	6.7	0.78		
- not too happy	4.0	1.18		
Taken all together, would you say you are?				
- very happy	9.3	0.90	1.1.1b	H13
- quite happy	7.2	0.40		
- not very happy	3.9	0.70		
- not at all happy	1.0	0.89		
Is your life at the moment?				
- very happy	9.3	0.90	1.1.1b	H14
- quite happy	7.2	0.40		
- quite unhappy	3.9	0.70		
- very unhappy	1.0	0.89		
How happy do you feel as you live now? Please choose one item from this card, that comes closest to your feeling				
- very happy	9.4	0.66	1.1.1c	H15
- fairly happy	6.9	0.54		
- neither happy or unhappy	5.1	0.30		
- fairly unhappy	3.4	0.49		
- very unhappy	0.6	0.80		

* Combined categories. Scores on separate categories not available in some cases

Exhibit 7/3.3b
Expert ratings on a 0 - 10 scale of responses to equivalent questions on happiness-in-life
Mean ratings by response category. N = 10

response category	*Mean rating in the context of question type:* *H1*	*H3 H2*	*H4*	*H6 H5*	*H7*	*H10 H8 H9*	*H13 H11 H12*	*H14*	*H18 H15 H17 H16*	*Overall mean*
very happy	9.2	9.2	9.3	9.2	9.0	9.3	9.3	9.4		9.3
quite happy							7.2			7.2
happy*							7.0		7.0	
pretty happy						6.7				6.7
fairly happy	6.4	6.3	6.5	6.3	6.8			6.9		6.5
rather happy*									6.0	6.0
neither happy nor unhappy								5.1		5.1
rather unhappy*									4.0	4.0
not too happy					4.2	4.0				4.1
not very happy	3.5		3.5				3.9			3.7
fairly unhappy								3.4		3.4
not happy				2.7						2.7
not very + not at all										2.5
not at all happy		1.1					1.0			1.0
very unhappy								0.6		0.6

* This item did not figure in the original weighting procedure and was estimated later
** Value assigned to combined categories where scores on separate categories were not available

The response category 'very happy' is used in different contexts: in three step scales, as well as in four step scales and five step scales. It is further used in contrast to 'not too happy' as well as 'not happy'. These variations hardly influence the ratings. The same applies to the often used category 'fairly' happy. Therefore we decided to use the average weights per response category irrespective of the question context. The averages are presented in exhibit 7/3.b.

Exhibit 7/3.3c
Expert ratings of responses on a 0 - 10 scale of response categories to equivalent questions about 'satisfaction with way-of-life'
Results by question. N = 10

Question	*Mean rating*	*Standard deviation*	*Location in part II* *table*	*question*
How satisfied are you with the way you are getting on now?				
- very satisfied	9.4	0.49	1.2.1a	S1+2+3
- all right	7.0	0.63		
- dissatisfied	2.9	1.04		
On the whole, how satisfied are you with the life you lead?				
- very satisfied	9.3	0.64	1.2.1b	S4+7
- fairly satisfied	6.5	0.50		
- not very satisfied	3.7	0.78		
- not at all satisfied	1.3	1.10		
On the whole, how satisfied are you with the life you lead?				
- fully satisfied	9.7	0.48	1.2.1b	S5
- not fully but to an extent satisfied	6.9	0.57		
- as yet unsatisfied	4.0	0.67		
- very unsatisfied	1.7	0.95		
On the whole, how satisfied are with the life you lead?				
- satisfied	8.5	0.71	1.2.1b	S6
- rather satisfied	6.8	0.42		
- as yet unsatisfied	4.0	0.67		
- unbearably dissatisfied	1.2	0.63		

7/3.3.3 Standardization by linear stretch

The above method of expert weighting of verbally labeled response categories is less appropriate if the difference between response scales is only the length. For instance in the case of the same question on life-satisfaction that is either scored on a 0-10 scale or on a 1-10 scale. In such cases simple linear transformation will do.

For happiness ratings, different scales are in use. Happiness is typically measured by self-report and cross-national studies on happiness mostly used single questions. An example of a commonly used item is presented below:

"Taking all together, how satisfied or dissatisfied are you currently with your life as a whole?"

1 2 3 4 5 6 7 8 9 10
Dissatisfied Satisfied

In this case, happiness is rated on a 10-step numerical scale. Other items use verbal rating scales, e.g., the 4-step rating scale

'very happy', 'fairly happy', 'not too happy' and 'unhappy'.

Happiness may be also rated on pictorial scales using smily and other graphical scales. Whatever the scale used, the respondent has to select one out of a limited number of discrete ratings, which is recorded eventually as a number, in the above scales one of the numbers from the sets {1(1)10} and {1(1)4} or e.g. {0(1)3} respectively.

For comparing results obtained by using different scales, the results of the primary numerical scale are subjected usually to a linear transformation onto a common 'secondary' scale. Below, we shall give the formulae to be used for this transformation.

Let r_1 = the rating on the primary scale,
h_1 = the rating on the primary scale for the most happy/satisfactory situation, and
u_1 = ditto for the most unhappy/unsatisfactory situation.

In the above first example $u_1 = 1$ and $h_1 = 10$.
The ratings after transformation will be denoted r_2, h_2 and u_2 respectively.
In most studies $h >$ u is chosen, so $u \leq r \leq h$. Some researchers, however, prefer $u > h$ and in the latter case $h \leq r \leq u$.

The three underlying assumptions for the linear transformation of happiness ratings are:

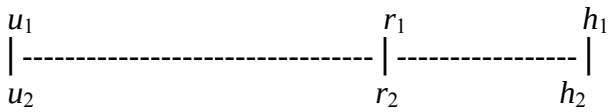
(a) the possible ratings of the primary scale can be considered as 'equidistant', so it is admissible to process the ratings as observations at the 'metric' level of measurement,

(b) $u_1 \rightarrow u_2$, and

(c) $h_1 \rightarrow h_2$.

The last two assumptions mean that the extreme possible ratings of the primary and the secondary scale are supposed to correspond perfectly to the same verbal or pictorial description label.

The situation in which $h_1 > u_1$ and $h_2 > u_2$ can be represented as follows:



From the proportionality

$$(r_1 - u_1)/(r_2 - u_2) = (h_1 - u_1)/(h_2 - u_2), \qquad (1)$$

it follows for the linear transformation, that

$$r_1 \rightarrow r_2 = u_2 + (r_1 - u_1)(h_2 - u_2)/(h_1 - u_1). \qquad (2)$$

As the reader can verify easily, this formula also holds in case $h_1 < u_1$ and/or $h_2 < u_2$.

The formula (2) can also be applied to the linear transformation of **mean values** m:

$$m_1 \rightarrow m_2 = u_2 + (m_1 - u_1)(h_2 - u_2)/(h_1 - u_1). \qquad (3)$$

For the corresponding **standard deviation** s, the transformation formula is

$$s_1 \rightarrow s_2 = s_1 \cdot |(h_2 - u_2)/(h_1 - u_1)|. \qquad (4)$$

This is based on the fact that, when x is a random variable and a and c are constants, then

$$\text{var}\,(ax+c) = a^2 \,\text{var}\,(x), \qquad (5)$$

so

$$\text{s}(ax+c) = a \cdot \text{s}(x). \qquad (6)$$

Example.
As an example we consider the transformation of $m_1 = 2.15$ and $s_1 = 0.64$ as the results of measurements obtained using the above 4-step rating scale

1 2 3 4
'very happy', 'fairly happy', 'not too happy', 'unhappy'.

We want to transform those statistics onto an 11-step scale with $u_2 = 0$ and $h_2 = 10$. This is the usual secondary scale in studies of happiness in nations. In that case the corresponding transformation formulae are

$$m_1 \rightarrow m_2 = 10\cdot(m_1-u_1)/(h_1-u_1) \quad (7)$$

and

$$s_1 \rightarrow s_2 = 10\cdot s_1/|h_1-u_1|. \quad (8)$$

Inserting $h_1 = 1$, $u_1 = 4$, $m_1 = 2.15$ and $s_1 = 0.64$ respectively results into the values $m_2 = 6.17$ and $s_2 = 2.13$ for the corresponding statistics on the [0;10] scale.

In this way we transformed average scores on 10 step-satisfaction in table 1.2.2a to an 11- step 0-10 score, which is presented next to the original means in table 1.2.2b.

Also some incidental scores were transformed linearly. In table 1.3 1-9 range scores from Israeli studies were upgraded to 0-10. In table 1.4 an Australian 1-9 rating is cut down to the common range 1-7.

7/3.3.4 Expert-rating and linear-stretch compared

One can of course wonder whether the latter method of linear transformation is preferable to the former method of expert ratings. Isn't that objective arithmetic rule preferable to subjective estimates by judges, and isn't the transformation by expert-ratings, essentially, also a correction for length of the rating-scales?

An evident objection is that linear transformation works only if the extremes of rating-scales represent the same 'true' happiness level and if the distances between successive steps are equal. These requirements are met only when rating-scales are graphical or numerically divided in equal steps, and when only the extremes are labeled verbally with identical words. These requirements are not quite met in most of the cases at hand here. Still, one could argue that the size of the difference between these methods is too small to take the trouble of making expert-ratings.

Therefore, we inspected whether the results of linear transformation differ substantially from the above discussed method of expert weighting. Both methods were applied to scores on questions on happiness-in-life (type O-HL) in six industrialized nations around 1980. The results are presented in exhibit 7/3.3.4.

The results differ considerably indeed. As we can see linear transformation produces higher scores on items with longer scales, whereas expert rating does not.

Exhibit 7/3.3.4
Two methods of transformation compared: Expert weighting and linear-transformation

country	*method*	*happiness question*		
		3-step (table 1.1.1a)	*4-step (table 1.1.1b)*	*5-step (table 1.1.1c)*
Australia	expert	7.4	7.8	7.6
	linear	6.5	7.4	8.3
Germany (W)	expert	6.6	6.9	6.2
	linear	5.1	6.5	6.8
France	expert	6.6	7.2	6.4
	linear	5.1	7.0	7.0
Italy	expert	5.8	6.5	6.1
	linear	3.7	6.1	6.6
Japan	expert	6.2	6.9	6.2
	linear	4.2	6.6	6.7
USA	expert	7.2	7.6	7.4
	linear	6.0	7.4	8.2

7/3.3.5 Validity of expert-rating

It is of course possible that the expert transformation involves a considerable distortion. The experts may have attributed weights to response categories that differ from the meanings respondents had in mind when answering the same questions. Therefore, the transformed scores may not provide a good estimate of true happiness in nations, in particular not of differences with that respect. The validity of our transformed scores can again be tested in two ways: by test for congruent validity and by test for concurrent validity.

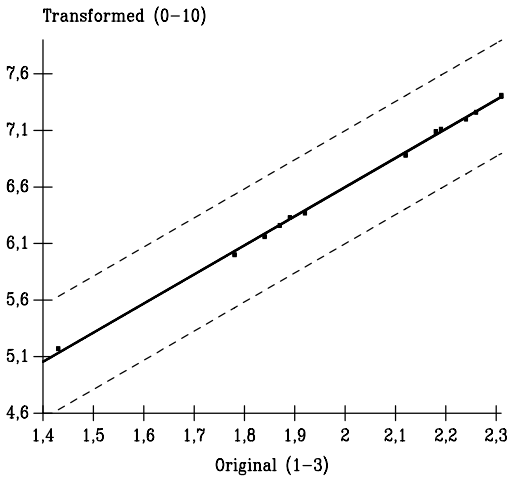
An evident test for congruent validity is assessing the correlation between transformed and untransformed scores. If we assume that the original scores provide a valid estimate of happiness in nations, a perfect correlation means that the transformed scores do equally well. If the correlation is not perfect, there are three possible explanations. The first possibility is then that the transformed scores estimate true happiness less well than the original scores (the above mentioned possibility). Reversibly, the second possibility is that the averages based on expert weighting are in fact closer to true happiness than the untransformed scores. This could be so, because the latter assume equal distances between rating options, whereas the expert ratings do not. Thirdly, it is of course possible that both are flawed in different ways. If transformed and untransformed scores are imperfectly correlated, further tests for external current validity must decide which is the best; for example by inspecting which variant explains most of the variance in quality of living conditions in nations, in an analysis as shown in section

5/1.2. If however, transformed and untransformed scores appear to be perfectly correlated, it is highly probable that both measure true happiness adequately. Further testing for concurrent validity is not useful in that latter case.

We checked congruent validity of our expert rating on two datasets: once more the Gallup/Kettering World Survey and on World Value Study I. Again we used the regression procedure and the 10% interval around the regression line. Exhibit 7/3.3.5 presents the scattergrams of transformed and original scores on the 3-step question on happiness-in-life (H8) in the Gallup/Kettering survey. The correlation is perfect (r=+.99). The scores are neatly on the regression line. Apparently the transformation procedure involved no distortion.

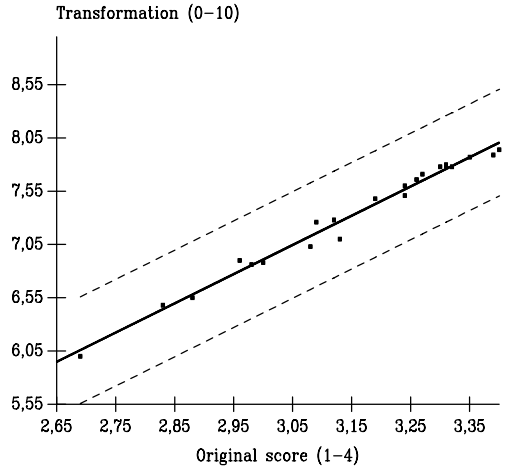
Exhibit 7/3.3.5
Transformed and original scores compared: 3-step and 4-step happiness

3 step happiness



r = +.99

4 step happiness



r = +.99

Data: Gallup/Kettering World Survey. See table 1.1.1a. World Value Study I. See table 1.1.1b.

We also considered the transformed and original scores on the 4-step happiness question in World Value Study I. Here again we see an almost perfect correlation (r=+.98) and all deviations are within the 10% interval.

We can conclude that our expert ratings successfully passed this validity test.

7/4 SUMMARY

This study collects the result of investigations that used acceptable measures of happiness. These acceptable measures are not quite identical however. This chapter explains how the divergent data were classified into equivalent categories. It further considers three techniques for transforming responses to dissimilar questions into comparable scores.

Grouping by focus
This database presents the data by kind of happiness measured. This breaks the data collection into four main parts: one big part on 'overall happiness' (code O), a smaller one on 'hedonic level' (code A) and two minor ones referring to 'contentment' (code c) and 'mixed indicators' (code M). Within these parts the collection is further differentiated in tables of near-identical indicators. This results in 13 tables with identical items and 5 tables for heterogeneous rest categories. Most of the tables with identical items concern overall happiness (10). Among these, three groups of questions some can be discerned which ask essentially the same thing, but that differ only in the rating of response. Though not 'identical', the items in these clusters are 'equivalent'. As such they qualify for conversion to a common scale. The possibilities for converting average scores on divergent indicators of happiness are however limited.

Transformation of scores on non-identical queries
Scores on indicators of different happiness variants can *not* be converted to the same standard. They measure essentially different things that do not necessarily coincide.

Non-equivalent queries Scores on different indicators of the same happiness variant can be converted in principle. However, in practice it is quite difficult to estimate the method effects involved. If sufficient data are available, we can inspect whether there is a linear relationship between responses yielded by different indicators in the same populations. Such data are only available for some single questions on overall happiness. We found a reliable relation in the nation scores on the two pairs of items: 1) 10-step life-satisfaction by 4-step satisfaction with way-of-life, and 2) 11-step life-satisfaction by 11-step best-worst possible life. In these cases missing values on one variable can be reliably estimated by linear regression on the basis of observed scores on the other; interpolation is less risky than extrapolation. In three pairs we found no reliable relation however: 1) happiness-in-life by satisfaction-with-life, 2) happiness-in-life by best-worst possible life, and 3) happiness-in-life by delighted-terrible life. In these latter cases we deem transformation inadvisable.

Equivalent queries Conversion is better possible when indicators (questions) are substantially equivalent, and differ only in number and labeling of response categories. In that case standardization by expert-weighting is justified. The expert-transformation applied here successfully passed a test for congruent validity.

If differences between equivalent items concern only the length of a graphic or numerical rating scale, simple linear transformation is most appropriate.

Only the latter two standardization methods (expert-weighting and linear

transformation) are applied in this data collection. In the tables transformed scores are mentioned for equivalent items. Transformed means are presented next to the original means.

Exhibit 7/3.3d
Expert ratings on a 0 - 10 scale of responses to equivalent questions type HAP 2.1
Mean ratings by response category N = 10

response category	*Mean rating in the context of question type:* *S1*	*S4* *S2* *S3*	*S5* *S7*	*S6*	*S8*	*Overall* *S9* *S10* *S11*	*mean*
fully satisfied				9.7			9.7
extra ordinary satisfied*						9.5	9.5
very satisfied		9.4	9.3			9.2	9.3
satisfied				8.5	8.5		8.5
all right	7.0						7.0
not fully but to some extent satisfied				6.9			6.9
rather satisfied*					6.8		6.8
pretty satisfied*		6.7					6.7
fairly satisfied			6.5			6.5	6.5
neither satisfied nor dissatisfied *					5.1		5.0
as yet unsatisfied			4.0	4.0	4.0		4.0
rather dissatisfied*							4.0
still dissatisfied*		4.0					4.0
not very satisfied		3.7			3.7		3.7
fairly dissatisfied*					3.4		3.3
dissatisfied		2.9				2.9	2.9
very unsatisfied				1.7			1.7
very dissatisfied						1.7	1.7
not at all satisfied		1.3					1.3
extremely dissatisfied*			1.2				1.2
unbearably dissatisfied					1.2		1.2

* These items did not figure in the original weighting procedure and were later estimated

REFERENCES

Andrews, F.M. & Withey, S.B (1976)
Social indicators of well-being,
New York, U.S.A., Plenum Press.

Gallup, G.H. (1976)
Human Needs and Satisfactions: A Global Survey
Public Opinion Quarterly, Vol.40, 1976, 459-467.

Gallup/Kettering (1976)
Human needs and satisfactions: A Global Survey. (Table books)
Charles F. Kettering Foundations and Gallup International Research Institute, USA

Michalos, A.(1991)
Global Report on Students Well-being, Volume I: Life Satisfaction and Happiness
New York, U.S.A., Springer-Verlag.

Torgerson, *(1958)*
Theory and methods of scaling
New York, U.S.A., Wiley.

World Value Surveys 1980-1993,
Cumulative file, 1994
ICPSR, AnnArbor USA, file nr 6160

This text is a revised version of chapter 7 in the book 'Happiness in Nations' by Ruut Veenhoven, RISBO, Erasmus University Rotterdam, 1993, Rotterdam Netherlands, ISBN 90-72597-46-X, pp.86-119