Social Indicators Research Series 68

Tineke de Jonge
Ruut Veenhoven
Wim Kalmijn

Diversity in Survey Questions on the Same Topic

Techniques for Improving Comparability

Springer

Social Indicators Research Series

Volume 68

Series Editor
Alex C. Michalos, Faculty of Arts Office, Brandon University, Brandon, Manitoba, Canada

Editors
Ed Diener, University of Illinois, Champaign, USA
Wolfgang Glatzer, J.W. Goethe University, Frankfurt am Main, Germany
Torbjorn Moum, University of Oslo, Norway
Mirjam A.G. Sprangers, University of Amsterdam, The Netherlands
Joachim Vogel, Central Bureau of Statistics, Stockholm, Sweden
Ruut Veenhoven, Erasmus University, Rotterdam, The Netherlands

This series aims to provide a public forum for single treatises and collections of papers on social indicators research that are too long to be published in our journal *Social Indicators Research*. Like the journal, the book series deals with statistical assessments of the quality of life from a broad perspective. It welcomes the research on a wide variety of substantive areas, including health, crime, housing, education, family life, leisure activities, transportation, mobility, economics, work, religion and environmental issues. These areas of research will focus on the impact of key issues such as health on the overall quality of life and vice versa. An international review board, consisting of Ruut Veenhoven, Joachim Vogel, Ed Diener, Torbjorn Moum, Mirjam A.G. Sprangers and Wolfgang Glatzer, will ensure the high quality of the series as a whole.

More information about this series at http://www.springer.com/series/6548

Tineke de Jonge • Ruut Veenhoven
Wim Kalmijn

Diversity in Survey Questions on the Same Topic

Techniques for Improving Comparability

Springer

Tineke de Jonge
Erasmus Happiness Economics
Research Organisation
Erasmus University Rotterdam
Rotterdam, The Netherlands

Ruut Veenhoven
Erasmus Happiness Economics
Research Organisation
Erasmus University Rotterdam
Rotterdam, The Netherlands

Wim Kalmijn (deceased)
Erasmus Happiness Economics
Research Organisation
Erasmus University Rotterdam
Rotterdam, The Netherlands

ISSN 1387-6570 ISSN 2215-0099 (electronic)
Social Indicators Research Series
ISBN 978-3-319-53260-8 ISBN 978-3-319-53261-5 (eBook)
DOI 10.1007/978-3-319-53261-5

Library of Congress Control Number: 2017933832

© Springer International Publishing AG 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Today, we are investing more in scientific research than ever before in human history, and a search in Google Scholar suggests that in 2015 some eight million scientific publications were produced. All this research is assumed to result in the 'accumulation of knowledge', thought to take place through an exchange of information in this 'academic forum'. Reality is more complicated than this. There is not one academic forum, but rather many different market places, and exchange of knowledge in these bazaars is considerably hindered by the different languages and currencies used. As a result, much of the gathered information gets lost in the dust of libraries. Obstacles to accumulation of knowledge differ across academic settings. In this book, we address one particular problem in one particular research field.

This book is about a problem in survey research: in this case, large-scale questionnaire studies among samples of the general population in countries. We focus on 'social surveys' that are about the quality of life in nations. Most modern countries conduct such surveys periodically, for example, the 'Understanding Society' study in the UK, the Level of Living survey in Sweden and the Public Opinion Survey on the Life of the Nation in Japan. Then, there are international survey programs, such as the Eurobarometer, the European Social Survey, the Gallup World Poll and the World Values Survey. Worldwide, some million people take part in such surveys every year.

Social surveys address similar themes, such as social position and opinions. Many of the topics are identical; all the surveys have questions about age, sex and education, and most also ask about income and ethnicity. Yet, the questions asked on these matters are often formulated differently, for example, questions about the education level of the respondent sometimes ask the respondent to give a number for years of education, while in other surveys, a list of education types is given, from which the respondent must select the final type achieved. In this case, many different classifications of school level can be used which vary within and between countries. In this mixture of ways of asking the same question of education levels

across nations and time, typically less than half of the available survey data can be used for such purpose.

One of the common topics in social surveys is 'happiness', that is, the satisfaction with one's life as a whole. This matter is typically measured using single direct questions, such as: 'Taking all together, how happy would you say you are these days? Would you say you are very happy, pretty happy or not too happy?' To date (2016), such questions have figured in some 10,000 survey studies and have been answered by some 200,000 respondents. The observed distributions of these responses are gathered in the collection 'Happiness in Nations' in the World Database of Happiness (WDH) (Veenhoven 2016). Viewing this collection of data, we can see that differently formulated questions have been used and that there are also many differences in the response options presented to respondents. Again, this means that only part of the data is available for comparing happiness across nations and time. These problems are not confined to comparing levels of education and happiness across populations and nations. They are found for many other topics in survey research, such as health care or customer satisfaction,

In this book, we present methods for dealing with this diversity in survey questions on the same subject; we review existing methods used to homogenize data and propose new ones. The book is a spin-off from the World Database of Happiness, the main aim of which is to collate and make available research findings on the subjective enjoyment of life and to prepare these data for research synthesis. The first methods we discuss were proposed in the book *Happiness in Nations* (Veenhoven 1993, Chap. 7 'How the Data Are Homogenized') which were used at the inception of the World Database of Happiness. Some 10 years later, a new method was introduced: the International Happiness Scale Interval Study (HSIS) (Veenhoven 2008). Taking the HSIS as a basis, Wim Kalmijn (2010) developed the Continuum Approach. Then, building on Kalmijn's work, Tineke DeJonge (2015) developed the Reference Distribution Method.

In this book, we describe the evolution of these innovations and provide a view of where we stand now. We go on to suggest ways to evolve this line of research. The book is based on the doctoral dissertation of Tineke DeJonge (2015); most chapters are based on articles authored in most cases by the three of us and Lidia Arends of the Erasmus University Rotterdam.

Wim Kalmijn passed away in November 2015 aged 81; he contributed much to the preparation of this book. We thank Willem Saris of the University Pompeu Fabra in Barcelona for his comments on the draft version and Miranda Aldham-Breary, senior volunteer of the WDH, for improving the English.

Rotterdam, The Netherlands Ruut Veenhoven

References

De Jonge, T. (2015). *Different survey questions on the same topic. How to make responses comparable?* Dissertation, Erasmus University Rotterdam, The Netherlands. Enschede: Ipskamp Drukkers. Available at http://repub.eur.nl/pub/78339

Kalmijn, W.M. (2010). *Quantification of happiness inequality*. Dissertation, Erasmus University Rotterdam, The Netherlands. Enschede: Ipskamp Drukkers. Available at http://repub.eur.nl/pub/21777

Veenhoven, R. (1993). *Happiness in nations, subjective appreciation of life in 56 nations, 1946–1992* (Studies in social-cultural transformation, 2). Risbo, Erasmus University Rotterdam, Netherlands. Available at: https://personal.eur.nl/veenhoven/Pub1990s/93b-con.html

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* (Social Indicator Research Series, Vol. 35, pp. 45–58). Dordrecht: Springer Press.

Veenhoven, R. (2016). *Happiness in nations*. World Database of Happiness, Erasmus University Rotterdam. Available at: http://worlddatabaseofhappiness.eur.nl/hap_nat/nat_fp.php?mode=1

Contents

Part I Comparing Responses to Different Survey Questions on the Same Topic: Problems and Conventional Solutions

1 Diversity in Survey Items and the Comparability Problem 3
Introduction 3
An Impression of the Diversity in Survey Items 4
The Incomparability of Time Series from Different Surveys and Different Time Periods 7
Conventional Scale Transformation Methods for Items from Different Surveys 10
Linear Stretch 10
Semantic Judgment of Fixed Word Value 12
Some Words on Methods to Correct for Response Heterogeneity at the Individual Level 12
The Need for Further Innovation 14
References 14

Part II Innovation 1: The Happiness Scale Interval Study

2 The Happiness Scale Interval Study 19
Introduction to the Happiness Scale Interval Study 19
The Scale Interval Recorder 19
Difference with Conventional Methods for Scale Transformation . . 21
Three Scale Transformation Methods Applied to Empirical Data 21
The Three Transformation Methods and the Comparability Problem 22
Discussion 24
References 25

3 Use of Happiness Scale Interval Studies in This Book 27
Research Questions Addressed in This Book Using HSIS-Results 27
HSIS-Studies Used and the Selection of Survey Items 28
The USA 28
The Netherlands 28
Spain 29
Representativeness of Judges 29
References 30

4 Equivalence of Rating Scales Using Different Keywords 31
Surveys with Equivalent Items for 'Happiness' and 'Satisfaction with Life' 31
The Problem 33
The Scale Interval Recorder as an Instrument to Compare the Intervals Represented by Equivalent Response Options 33
Differences in Value Ranges Represented by Response Options Labeled in Dutch, Spanish and English 34
Response Options Labeled in Dutch 34
Response Options Labeled in Spanish 38
Response Options Labeled in English 41
Discussion 44
Methodological Consideration 44
Limitations 44
Implication of the Method 45
Advice for Further Research 45
Conclusion 45
References 46

5 'Very Happy' Is Not Always Equally Happy 47
Evaluation of the Labels of Response Options in the Context of the Scale 47
The Semantic Judgment of Fixed Word Value Method Revisited 47
The Context of the Scale and Comparability 48
Interpretation of Response Options of the Same Item in Different Languages 50
Translation and Comparability 50
Difference in Interpretation Between Different Languages 51
Different Questions About the Same Topic, But with Identical Response Options 54
The Leading Question and Comparability 54
Difference in Interpretation Given a Difference in Leading Question 55

Discussion . . . 58
Conclusion . . . 59
References . . . 60

6 The Labeling of Anchor Points and the Occurrence of Zero-Width Intervals . . . 61
Anchor Points and Zero-Width Intervals . . . 61
The Occurrence of Zero-Width Intervals . . . 62
Discussion . . . 66
Conclusion . . . 67
References . . . 67

Part III Innovation 2: The Continuum Approach

7 The Continuum Approach . . . 71
Happiness: A Discretely or Continuously Distributed Variable? . . . 71
Outline of the Continuum Approach Applied to Subjective Well-Being . . . 72
Combination of the Continuum Approach with the Scale Interval Method . . . 74
The Continuum Approach and Discrete Numerical Scales . . . 76
Comparison of the Estimated Means Using Different Methods . . . 78
References . . . 80

Part IV Innovation 3: The Reference Distribution Method

8 The Reference Distribution Method . . . 85
Using a Reference Distribution to Derive Boundaries Between Response Options . . . 85
Illustration of Applying the Reference Distribution Method . . . 86
Scale Transformation Using the Reference Distribution Method . . . 89
Application of the Reference Distribution Method . . . 91
Discussion . . . 94
Strengths of Scale Transformation Using a Reference Distribution . . . 95
Limitations . . . 95
Conclusion . . . 96
References . . . 96

9 Stability of the Boundaries Between Response Options for Different Countries 97
Research Question 97
Approach for Testing the Stability of Boundaries 99
Measurements Using the ESS and EB Items from 2012 100
Frequencies Distributions of the EB Time Series 104
Deviation from the Reference Boundaries in Horizontal Direction 106
Deviation from the Beta Distribution in Vertical Direction 109
Discussion 113
Conclusion 113
References 113

10 Analysis of Differences in Trends Among the Satisfied Few and the Dissatisfied Few 115
Mean Satisfaction in Subgroups of the Population 115
Differences in Trends Among the Satisfied Few and the Dissatisfied Few 116
Discussion 120
Reference 121

11 Combining and Pooling of Time Series on Life Satisfaction in the USA, Japan, The Netherlands and Spain 123
Research Question 123
Distortion of Trends Due to Biases in Measurement 124
Survey Mode Effects 124
Ordering of Questions 125
Response Shift 126
Contextual Influences 127
Time Series on Life Satisfaction in the USA, Japan, The Netherlands and Spain 128
Searching for Sources for Response Bias to Correct 130
Preparation for the Conversion of Time Series 132
Combining Converted Survey Results into Long Time Series 138
Conversion of Population Means for Time Series of Individual Survey Items 139
Pooling of Converted Time Series 143
Discussion 144
Conclusion 145
References 146

12 Directions for Further Research 149
Refinements 149
Differential Interpretation of Items by Subgroups 149
Improvement of HSIS Ratings 150
Improvement of Estimates Obtained with the Reference Distribution Method 150
Comparison of the Estimates Obtained with the HSIS and the Reference Distribution Method 152
Applications 152
Application in Research Synthesis of Happiness 152
Application in New Research on Happiness 153
Application to Other Topics Than Happiness 154
References 155

Appendices 157
Appendix A: Original and English Wording of Survey Items on Happiness from the Non-English HSIS-Studies Used in This Book . . . 157
Appendix B: Statistics of the Assessments by Students in the HSIS-Studies Used 162
Appendix C: Cumulative Frequencies and Parameters Beta Distributions for Groups of European Countries 175
Appendix D: Horizontal and Vertical Deviation of Reference Boundaries for Groups of European Countries 178
Appendix E: Cumulative Distributions and Parameters Beta Distribution 183
E.1: Cumulative Frequencies and Parameters Beta Distribution USA 183
E.2: Cumulative Frequencies and Parameters Beta Distribution Japan 187
E.3: Cumulative Frequencies and Parameters Beta Distribution The Netherlands 191
E.4: Cumulative Frequencies and Parameters Beta Distribution Spain 197

Glossary 201

Abbreviations

BRFSS	Behavioral Risk Factor Surveillance System
CAPI	Computer-assisted personal interviewing
CBS	Statistics Netherlands
DHS	Dutch Household Survey
EAVS	East Asia Value Survey
EB	Eurobarometer
EDAC	European Data Center for Work and Welfare
EQLS	European Quality of Life Survey
ESS	European Social Survey
FLS	Future Life Survey
HSIS	Happiness Scale Interval Study
ISSP	International Periodical Social Survey Program
LAPOP	Latin American Public Opinion Project
LDC	Leisure Development Centre
LIN	Life-in-Nation Survey
LPS	Lifestyle Preference Survey
NIPO	Dutch Institute for Public Opinion
OECD	Organization for Economic Co-operation and Development
%SM	Percentage of scale maximum
SCP	The Netherlands Institute for Social Research
SWB	Subjective well-being
TRAPD	Translation, Review, Adjudication, Pretesting and Documentation
VAS	Visual analogue scale
WDH	World Database of Happiness
WVS	World Values Survey

Part I
Comparing Responses to Different Survey Questions on the Same Topic: Problems and Conventional Solutions

Chapter 1
Diversity in Survey Items and the Comparability Problem

Introduction

Survey research is a major method used in the social sciences and is largely based on standard questions with pre-coded response options called 'response scales' to which respondents answer by picking one of the options. There is little uniformity in the survey items[1] used. This difference in items is no problem when surveys are analysed separately, but it limits the comparability of findings gathered in different surveys that used different items for the same topic. This reduces the value of our accumulating of knowledge and calls for techniques to improve the comparability of data.

This diversity in the wordings of questions and in response options also appears in survey research on subjective well-being which took off in the 1970s in the wake of the Social Indicator Revolution. In this context Campbell et al. (1976) set up a program for monitoring the quality of American life, with emphasize on the experience of life rather than the, mostly material, conditions of life thus shifting the focus from current economic goals to subjective well-being. They recognized that they would encounter many problems of definitions and methods as they could not fall back on standardized measures established in earlier research which would fulfill their needs. Their main focus was upon expressions of satisfaction with specific domains of life experience, but they also included an item on happiness and an item on satisfaction with life as a whole in their survey. Andrews and Withey (1976), other pioneers in this field of research, explored a large set of survey items, among which were questions on the subjective appreciation of one's life as a whole. Many more questions designed to measure subjective well-being have been used since then. To date about a 1,000 different questions on the subjective appreciation of one's life as a whole from some 10,000 studies have been gathered in the collection 'Measures of Happiness' of the World Database of Happiness

[1] We use the term 'item' for a survey question and its corresponding response options.

© Springer International Publishing AG 2017

T. de Jonge et al., *Diversity in Survey Questions on the Same Topic*, Social Indicators Research Series 68, DOI 10.1007/978-3-319-53261-5_1

(Veenhoven 2016a, b). About half of the differences in questions are in the number and wording of response options, other differences are due to causes such as the reference to time and the visual presentation of the scale or the method of assessment. The World Database of Happiness (WDH) focuses on happiness in the sense of the subjective enjoyment of one's life as a whole (Veenhoven 1984). In this definition 'happiness' is synonymous with 'life satisfaction'. This concept of happiness is currently the one most commonly used in the social sciences and it lies at the heart of the WDH (Veenhoven 2011).

One of the aims of happiness researchers is to assess differences in happiness across nations. This requires comparison of data drawn from different surveys containing questions about happiness. In surveys however, different kinds of response scales are used, both verbal scales and numerical scales, and these scales also differ in the number of response options available, some including only two options, for example yes or no, and others as many as 11, for example 0–10 numerical scales. As a consequence only a part of the available research can be used by scholars in any comparison or meta studies. Likewise, another aim of happiness researchers is to compare happiness within countries over time. This also requires equivalent questions and response scales, but since the response scales used can change over the years, the number of comparable data available will often be inadequate for a valid comparison to be made.

Diversity in survey items is often treated in one of two ways: one is to abstain from any comparison when rating scales are not fully identical. This means that most of the findings on happiness are incomparable and thus lost for synthetic analysis. The other approach is to ignore the problem, typically by assuming that the ranks of the response options reflect the degree of happiness denoted and treating these numbers as metric values that can be transformed to the same range. This approach carries the danger of producing nonsense.

An Impression of the Diversity in Survey Items

Most people have a positive perception of their own well-being, at least in the western world. As a result, the distribution of responses to questions on happiness is skewed, with a long tail on the left that represents 'negative' outcomes (Campbell et al. 1976; Lee et al. 1982; Diener and Diener 1996; Cummins 2003; Frijters et al. Shields 2008; Guven et al. 2011). Irrespective of the scale used, this skewness has to be kept in mind when interpreting the results of such measurements.

Within the large set of existing measures of happiness, the number of response options and the distinction between verbal and numerical response scales are obvious variations. To meet the skewness of the distribution, in the past verbal scales were devised that are asymmetric due to mainly positively formulated response options. An example of such a scale is the one used by Statistics Netherlands from 1974 to 2010, a long period, to measure satisfaction with life. Respondents were given the response options 'Extraordinarily satisfied', 'Very satisfied',

'Satisfied', 'Fairly satisfied' and 'Not very satisfied'. Note: only the latter of these responses is formulated negatively. The idea behind this rather asymmetric scale at the time it was devised was that it would give the possibility for more variation in the responses than if a more symmetric scale was used. Another, totally different example of a response scale devised to reduce the skew is provided by Andrews and Withey (1976). They argued that substantially skewed distributions pose problems in analysis and that reducing the skew in the distribution of response would therefore enhance our ability to find meaningful relationships between different aspects of life. With this in mind they felt they could improve the "satisfaction" scale used by Campbell et al. which is a symmetric numerical scale, with one anchor point[2] labeled 'Completely satisfied' and the other anchor point labeled 'Completely dissatisfied'. Andrews and Withey's improvement of this scale (1976, pp. 18–19) consisted of adding more affect to it, resulting in what is known as the Delighted-Terrible scale. This scale includes the seven on-scale options 'Delighted', 'Pleased', 'Mostly satisfied', 'Mixed – about equally satisfied and dissatisfied', 'Mostly dissatisfied', 'Unhappy' and 'Terrible', supplemented with three off-scale options 'Neutral – neither satisfied nor dissatisfied', 'Does not apply to me' and 'I never thought about it'. Andrews and Withey drawing conclusions from the numerous different ways they experimented with to measuring affect, stated they believed that this scale was the most effective.

The satisfaction with life scale used by Statistics Netherlands is a unipolar scale: all response options contain the word 'satisfied'. This differs from a bipolar scale, as for example the "Satisfaction with life scale" of Campbell et al. where in the response options, for example, the word 'dissatisfied' is used as the opponent of 'satisfied'. Furthermore, a scale does not necessarily need to have a neutral mid-point dividing it into a positive and a negative pole and the end points of different scales may vary in the extremity of the wording used, for example 'extraordinarily' is more extreme than 'very' but both are subject to a respondents interpretation of the words the effect of the context in which they are used, and this will vary from respondent to respondent.

Most of the variations discussed above hold for both verbal response scales and for numerical response scales. Although numbers are used on a numerical scale to express the respondent's degree of happiness, it is still necessary to use words to describe what the anchor points of the scale denote, and it is this wording which defines whether the scale is conceived to be unipolar or bipolar. The wording of such descriptions can include the subject of measurement, as in 'dissatisfied' – 'satisfied' or leave it to the respondents how they interpret the anchor points or extremes of the scale when a formulation is given in terms like 'best possible' – 'worst possible'. An early and well-known example of the latter is Cantril's self-anchoring Ladder of Life Scale. Using this satisfaction of life measure, respondents

[2]We use the term 'anchor points' for the response options at both ends of a discrete scale. In the case of a continuous distribution, we use the term 'extremes' to refer to the boundaries of the continuum that bounds this distribution.

are asked to imagine a ladder with steps numbered from zero at the bottom to ten at the top. The top step of the ladder represents the best possible life for the respondent and the bottom step the worst possible life. The respondents have to rate on which step of the ladder they feel they personally stand at the time of questioning. This measure was first employed for a national sample in the United States in 1959 (Campbell et al. 1976, p. 31) and is still used, for example in Gallup's World Poll (Bjørnskov 2010; Gallup 2016).

Other variations in numerical scales are the visual orientation, which can be vertical or horizontal, and the labeling of the anchor points that can go from negative to positive, for example −5 to +5, consists only of positive numbers and possibly including zero starting at 0 or 1, or there can be no numbering (Mazaheri and Theuns 2009). In an experiment done by Schwartz et al. (1991) using an 11-point numerical scale with anchor points labeled from 'Not at all successful' to 'Extremely successful' and ranging from −5 to +5 only 13% of the respondents gave an answer between −5 and 0. When the range changed from 0 to 10, the percentage of answers at the lower end of the scale changed to 34%. A similar result was found by Sangster et al. (2001). From this experiment Schwarz and his colleagues concluded that a numerical scale starting at zero suggests the absence or presence of the subject under study, which makes the scale unipolar. If conversely, one half of the scale is negative and the other half is positive, then the positive values are related to the presence of the subject one is interested in, whereas the negative values represent the opposite. Schwartz et al. also suggest that scales that are intended to assess the intensity of a single attribute, for example happiness, should follow a zero-to-positive-value format to emphasize that the question pertains to the absence or presence of this specific attribute, rather than the presence of its opposite. This suggestion is underpinned in an elaborated discussion on happiness as a variable in Kalmijn (2010, Ch. 2). In his thesis, Kalmijn devotes a section to our perspectives on the nature of happiness and satisfaction, the difference between intensity and extensity variables, the polarity of happiness scales and the level of measurement.

The response scale cannot be seen separately from the related leading survey question. Variations in the wording of a question also lead to numerous different survey items. Furthermore, the time frame a question relates to leads to more variations. For example, a question can refer to satisfaction with life over the life time or just at the moment of questioning or for the last 4 weeks. Moreover, a question can contain a keyword such as the word 'happy' in the question "Are you happy with your life?", where the subject can either be explicitly formulated in the response options or not. Likewise a question can be formulated as "Do you feel . . .?" with a keyword only mentioned in the labels of the response options of the related scale.

These are just some examples of the variations in the wording of the questions used in happiness research. Of course there are many more variations one can think of. Each variation will influence the response patterns to a question and can induce structural breaks in the data that can change results dramatically (Cummins and Gullone 2000; Bjørnskov 2010, p. 43; Pudney 2010, p. 6). According to Bjørnskov,

Table 1.1 Classification of survey questions on happiness in the world database of happiness

Aspect	Example	Code
Keyword used	Satisfaction with life	O-SL
Time reference	Currently	c
Method of assessment	Single question	sq
Kind of rating scale	Verbal	v
Length of rating scale	4-step	4
Variant of rating scale	Agree – disagree	a, b,... etc.

a fundamental discussion on this problem, including the framing of life satisfaction questions, has been going on since the 1940s. A comprehensive description of the variations in items and a discussion on these variations is given in, among others, Saris and Gallhofer (2007) and Mazaheri and Theuns (2009). Additionally, a systematic overview of all the variations in survey items on happiness can be found in the collection 'Measures of Happiness' of the WDH. The measures in the WDH are classified by six aspects, see Table 1.1, and the survey questions presented in this book are coded according to this classification.

The Incomparability of Time Series from Different Surveys and Different Time Periods

The Social Indicator Revolution ushered in an era of periodic measurements of several social indicators, among which that for life satisfaction. We will illustrate the diversity in survey items and the comparability problem by using time series on life satisfaction collected in the United States, Japan and The Netherlands. References to these time series and the items used for measuring life satisfaction can be found in the WDH (Veenhoven 2016b).

United States The Gallup organization started measuring life satisfaction in the United States in its Gallup Poll[3] periodically in 1973, asking "In general, are you satisfied or dissatisfied with the way things are going in your own personal life?" using a 2-point verbal scale with the response options labeled 'Satisfied' and 'Dissatisfied'. These same response options were used in the item with the slightly different question "In general, are you satisfied with the way things are going in your personal life at this time?" which Gallup introduced in 2001. At the same time Gallup introduced a second variation on the 1973-item, consisting of the question "In general, how satisfied or dissatisfied are you with the way things are going in your personal life at this time?" and a 4-point verbal scale with the response options labeled 'Very satisfied', 'Somewhat satisfied', 'Somewhat dissatisfied' and 'Very dissatisfied'. These response options were also used in the third variation coming

[3] http://www.gallup.com/products/170987/gallup-analytics.aspx. Assessed 3 February 2016.

with the question "Overall, how satisfied are you with your life…?" which was included in the Gallup Poll in 2005.

Japan In Japan the periodic measurement of life satisfaction started even earlier than in the USA, using a 4-point verbal scale item which was included in the Life-in-Nation survey in 1958. This item consisted of the question "By the way, how do you feel about your life?" and response options labeled 'My life could be better, but on the whole I am satisfied with my current life', 'I am not satisfied with my current life, but it is not too bad to keep more or less on the current level', 'My current life is far from satisfactory', 'I cannot stand my current life'. This item was replaced by a new item in 1964, consisting of the question "How do you feel about your life?" and the response options 'Sufficiently satisfied', 'Rather satisfied, but not sufficiently', 'Fairly dissatisfied' and 'Extremely dissatisfied'. In its turn, this item was replaced in 1970 by the item consisting of the question "How do you feel with life these days" and the response options 'Very satisfied', 'Somewhat satisfied', 'Somewhat dissatisfied' and 'Very dissatisfied'. A replacement of the life satisfaction item in the Life-in-Nation survey followed in 1992, the current item consists of the question "Overall, to what degree are you satisfied with your life these days?" with the response options 'Satisfied', 'Fairly satisfied', 'Somewhat dissatisfied', 'Dissatisfied'.

The Netherlands One of the eldest time series on life satisfaction in The Netherlands comes from a series of measurements by Statistics Netherlands (CBS), which started in 1974 with the launch of the first Life Situation Survey which CBS developed at the request of, and in close collaboration with, The Netherlands Institute for Social Research (SCP). The item used for these measurements was the 5-point verbal scale item consisting of the question "To what extent are you satisfied with the life you currently" with the response options 'Extraordinarily satisfied', 'Very satisfied', 'Satisfied', 'Fairly satisfied' and 'Not very satisfied'. This item was used in changing surveys by both SCP and CBS with different periodicities over a period of almost 40 years (DeJonge 2009): CBS used the item for life satisfaction until 2010. After having conducted a split-half experiment in 2012, in which a verbal and a numerical scale were used to measure life satisfaction, CBS decided to change to a 10-point numerical response scale (Van Beuningen et al. 2014). The SCP used the verbal scale until 2002 when it changed to a 10-point numerical scale for life satisfaction.

The Rank Method In survey research it is common practice to assign ranks to the response options of a discrete scale to calculate a sample mean, regardless of the semantics of the wording used to label the options. The sample mean is accordingly calculated as the weighted average of the ranks of the response options using the relative frequencies as weights. In this common practice, denoted the Rank Method, it is implicitly assumed that equivalent response options in equivalent scales are evaluated identically and that the response options are equally distanced regardless of the topic of concern. The value range assigned to the words by which a response option is labeled, however, heavily depends on the context of the scale as does the