

Measurement Scales for Scoring or Ranking Sets of Interrelated Items

Luigi Fabbri

Abstract Surveys concerned with human values, economic utilities, organisational features, customer or citizen satisfaction, or with preferences or choices among a set of items may aim at estimating either a ranking or a scoring of the choice set. In this paper we discuss the statistical and practical properties of five different techniques for data collection of a set of interrelated items; namely the ranking of items, the technique of picking the best/worst item, the partitioning of a fixed total among the items, the rating of each item and the paired comparison of all distinct pairs of items. Then, we discuss the feasibility of the use of each technique if a computer-assisted data-collection mode (e.g. CATI (telephone), CAPI (face-to-face), CAWI (web) or CASI (self-administered)) is adopted. The paper concludes with suggestions for the use of either technique in real survey contexts.

1 Introduction

In the following, we deal with survey procedures for the collection of data on a closed set of interrelated entities. The procedures apply to research situations in which questionnaires are administered to a population. The aim of these procedures is to elicit choices or preferences. For the sake of clarity, a choice applies to expressions such “*I choose that option*”, while a preference applies to “*I prefer this instead of that*”. In either case, a choice/preference criterion is to be specified.

In marketing studies, these procedures are called “stated preferences”, as opposed to “revealed preferences” that stem from the observation of purchase data. The stated preference data aim at describing the world (market, society, community, etc.) as it could be; the revealed preference data represent the world as it is.

Luigi Fabbri (✉)
Statistics Department, University of Padua, Padua, Italy
e-mail: luigi.fabbri@unipd.it

It is assumed that the choice set is composed of p entities and that data are collected from n sample units. The data can be used to estimate either scores or ranks referred to as the choice set. Scores may be of interest to parameterise a function; for instance, to compose a complex indicator by combining a set of p elementary indicators for the final aim of partitioning resources among the units of a population. Ranks are appropriate, for instance, to define priorities among the choice set, to help communicators to make the public aware of hierarchies among the items, or to highlight the units that gained or lost positions with respect to a previous ranking.

As an example, if a budget is to be partitioned among the best-performing universities, quantitative and precise values are needed. If, instead, the best-performing universities are to be listed for guiding students who are choosing a university, the accuracy of the three top positions may suffice.

The data we are concerned with are normally administered as a battery of items under a common heading. The schemes of item presentation will be described in Sect. 2. We will also evaluate the procedures from the statistical (Sect. 3) and practical (Sect. 4) viewpoints with the aim of using either procedure in a computer-assisted survey. In Sect. 5, some general conclusions will be presented.

2 Survey Procedures

To collect choice or preference data, a battery of questions is to be administered to a sample of n ($n \geq 1$) individuals. Let us suppose that the choice set is selected from the universe of all possible sets that satisfy certain statistical properties. The following techniques will be examined: ranking (Sect. 2.1), picking the best/worst item (Sect. 2.2), partitioning a fixed budget (Sect. 2.3), rating (Sect. 2.4) and comparing in pairs (Sect. 2.5). An example of how a choice set can be constructed for a conjoint analysis¹ will be presented, as well (Sect. 2.6).

2.1 Ranking

The final result of the application of a ranking technique to a sample of respondents is a sequence from the most relevant item to last, according to an underlying construct. The procedure requires that each respondent evaluate all the items of the set simultaneously. Figure 1 is a scheme of a typical question for ranking purposes.

The technique may be administered in other formats whose common purpose is to determine a hierarchy of the p items, normally without ties. Ties may be allowed if the researcher perceives that the task required is too difficult for respondents. An alternative procedure that generates a ranking is the request of a sequence of choices

¹ Conjoint analysis is a multivariate method of statistical analysis of a set of alternatives each of which is qualified by two or more joint categories.

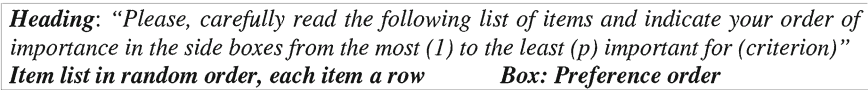


Fig. 1 Scheme of a question for ranking a choice set

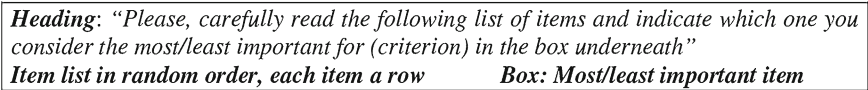


Fig. 2 Scheme of a question on picking the best/worst item

of the kind “best-worst”: initially, the choice of the best and worst items is performed using all p -listed alternatives, the second using the $p - 2$ remaining alternatives, and so on with two divergent choices at a time until the list exhausts. Munson and McIntyre (1979) and McCarty and Shrum (1997, 2000) call this stepwise ranking “sequential most-least procedure”.

2.2 Picking the Best/Worst Item

If ranking is not possible in full, each respondent may be asked to pick the most valued item and in case the least valued item on the list. Other names for this technique include *most/least* and *max difference analysis*. Figure 2 is a scheme of a typical question for obtaining this objective.

Respondents are sometimes asked to pick k ($k < p$) items as the most/least preferable ones and, in some instances, to select the best/worst item among the k . In other cases, the request could be just to select the k most/least preferable items out of the choice set. This more complex procedure will be named “pick k out of p items”. Even the method of picking one or more items requires the simultaneous contrast of the whole set of entities, but the procedure is much less demanding for respondents than full ranking because just a few items are pre-selected for top or bottom positions after a first reading of the list.

2.3 Fixed Total Partitioning

A respondent is assigned a fixed budget of, let’s say, 100 points, and is asked to partition it over the p items according to some criterion. This procedure is also named “budget partitioning”. Figure 3 is a scheme of a question for fixed total partitioning.

If a computer-assisted interviewing system is applied, it is possible to check the sum of points spent after each assignment. Conrad et al. (2005) found that running

Heading: “Please, consider a budget of 100 points to be partitioned among the following items. Indicate in the side boxes how many points you would assign to each item according to (importance criterion)”	
Item list in random order, each item a row	Box: Assigned points

Fig. 3 Scheme of a question for fixed total partitioning

Heading: “Please, assign an importance score between 1 and c, where 1 is the minimum and c the maximum, to each of the following items”	
Item list in random order, each item a row	Minimum=① ② ③ ④=Maximum

Fig. 4 Scheme of a question on rating individual items

totals improved the likelihood that the final tally matched the given sum, led to more edits of the individual items (the Authors considered this occurrence a symptom of the importance attributed by respondents to the required task) and took the respondents less time. A researcher can decide whether to insist that the respondent adjusts the assigned points if their sum does not fit the fixed total. Each item i ($i = 1, \dots, p$) is assigned by a respondent a “quantum” of a criterion, e.g. importance, relatively to the overall budget. It is relevant for survey purposes that respondents can assign no points at all to unimportant items, a possibility that is excluded if the rating technique is applied.

Any point assignment is an interval score that can be contrasted with the scores of the other $(p - 1)$ items and the scores given by other respondents to item i . A researcher who does not rely on the obtained values may retain just the ordinal hierarchies between the items, transforming in this way the data into a ranking.

2.4 Rating Individual Items

The respondent is asked to assign each item i ($i = 1, \dots, p$) an independent value according to either an ordinal (Likert-type) or an interval scale. The measurement scale is the same for all items. In general, in each row, there is the item description and a common scale with the so-called anchor points (minimum and maximum) at the extremes (see Fig. 4). With reference to weights estimation, the technique is sometime called “absolute measurement”.

A possible improvement to the simple rating technique is the combination of two methods; for instance, that of picking the most and the least important items and then rating all the items in the usual way (Krosnick and Alwin 1988) or to rank-and-rate as suggested by McCarty and Shrum (2000).

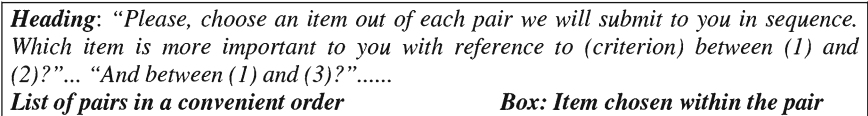


Fig. 5 Scheme of the administration of $p(p - 1)/2$ paired comparisons

2.5 Paired Comparisons

Each respondent is administered all the $p(p - 1)/2$ pairs of items and asked to choose the preferable item of any pair according to a predefined criterion. Ties may be admitted to avoid forcing reluctant respondents. Figure 5 is a scheme of question for paired comparisons.

Saaty (1977) proposes a variant of the full pair comparison technique. His technique involves submitting to a single volunteer ($n = 1$) the whole set of the $p(p - 1)/2$ pairs asking him or her not only which is the best item for each pair (i, j) but also how many times item i is more preferable than item j .

In addition, Clemans (1966) proposes a paired comparison system that is more complex than the basic one. It involves assigning a respondent 3 points for every two items of a pair so that the possible outcomes are (3, 0), (2, 1), (1, 2), (0, 3), yielding for each comparison a more differentiated outcome than the simple choice between the paired items.

If p is large, any paired comparison system is not viable even if visual aids and interviewers are invested. For this reason a reduced form of preference elicitation may be adopted that limits response burden. One of these techniques is the *tournament* technique as proposed by Fabbris and Fabris (2003). This involves (i) ordering the set of p items according to a criterion and (ii) submitting for choice in a hierarchical fashion the $p/2$ pairs of adjacent items, then the $p/4$ pairs of items preferred at the first level of choice, and so on until the most preferred item is sorted out. The choice may be structured by composing pairs or triplets of items at all levels (for more details see Sect. 4.5).

2.6 Conjoint Attribute Measurement

All the procedures presented in the previous sections apply for a conjoint analysis of sets of attributes. Conjoint measurement requires the definition of a set of alternatives that respondents have to rank, pick the best, rate or compare in pairs. Each alternative is specified as a set of attributes belonging to variables common to all alternatives. For instance, a conjoint measurement for market research purposes may be set up as in Fig. 6.

The four alternatives, also called “prop cards” (Green et al. 2001), are the logical product of two variables with two categories each and a related price/cost. Imagine,

<i>Alternative A</i>	<i>Alternative B</i>	<i>Alternative C</i>	<i>Alternative D</i>
$V_1=1$	$V_1=1$	$V_1=2$	$V_1=2$
$V_2=1$	$V_2=2$	$V_2=1$	$V_2=2$
Price: level 1	Price: level 2	Price: level 3	Price: level 4

Fig. 6 Example of items for a conjoint measurement

for instance, the transportation alternatives for reaching the university campus from a nearby town: the V_i s denote variables that might influence the students' choice (e.g. travel time, schedule, comfort and so on), and $V_i = j$ means that option j of variable V_i is specified.

The alternatives are presented in a simplified way, often illustrated by a photo or other visual representation that helps respondents to associate each alternative to its price, consciously contrasting the offered alternatives and evaluating the trade-offs between alternatives (Gustafsson et al. 2007). Conjoint analysis assumes that alternatives compensate each other, meaning that a trade-off exists between alternatives. This may be inadequate in the case of immoral or dangerous alternatives; for instance, an alternative may be a fatal level of pollution. Hence, the researcher should ignore alternatives that may be dangerous over a certain threshold.

Conjoint measurement can compare several alternatives. Even if the alternatives are qualified by variables whose number of categories to be crossed is large, the presentation can be simplified by planning an experiment that drastically reduces the number of cross-categories. Readers interested in this type of experiments could consult, among others, Green and Srinivasan (1990), Elrod et al. (1993), Kuhfeld et al. (1994) and Louviere et al. (2000).

3 Statistical Properties of the Scales

Let us consider the aim of estimating the population values of the p items on the basis of the preferences x_{hi} expressed by individual h ($h = 1, \dots, n$) for item i ($i = 1, \dots, p$). A function $g(X_i)$ links the possible responses to the latent value ξ_i of item i :

$$g(X_i) = \xi_i + \varepsilon_i \quad i = (1, \dots, p), \quad (1)$$

where ε_i is an error term that accounts for the sampling error, c_i , and the observational error, r_i , inherent to item i :

$$\varepsilon_i = c_i + r_i \quad i = (1, \dots, p). \quad (2)$$

Sampling and response errors are independent of each other and can be evaluated separately. The sampling error depends on the sampling design, and its evaluation

is beyond the scope of this paper.² The response error is composed of systematic (“bias”) and random components. The systematic component is common to all respondents, is constant regardless of the number of obtained responses and is at risk of being confused with the latent value ξ_i . The random errors vary within and between individuals. Response errors vary according to the following:

- *The respondent’s answering style*, e.g. his or her capacity and willingness to respond correctly (Baumgartner and Steenkamp 2001). The individual error may be either random (zero mean in repeated surveys) or systematic (constant over repeated surveys).
- *Group response biases*, i.e. common to a population segment or common to the whole assignment of an interviewer or supervisor. They are measurable as within-group correlated response errors (Harzing et al. 2009).
- *The survey context*, i.e. the set of social and physical characteristics of the interview setting. A non-neutral setting may generate systematic errors (Swait and Adamowicz 1996).

The expected value of the random error is nil across repeated measurements, but if the response error is correlated within groups of respondents, it may have a negative impact on the response variance of estimates (Fellegi 1964; Bassi and Fabbris 1997). Group response biases and interviewer biases belong to the family of intra-class correlated errors.³

The latent value ξ_i is the population value of item i that represents an aspect of the concept underlying all the items of a set. Values may relate to human life principles (Rokeach 1967, 1979); consumers or citizens’ preferences (Beatty et al. 1985); economic utility of facts, goods or services (Small and Rosen 1981); criteria for organizational decisions (Crosby et al. 1990; Aloysius et al. 2006) and so on. In econometric theory, ξ_i is the utility associated with the i th alternative; in consumer preference theory it is the individual part-worth parameter of an aspect related to a common construct; in multi-criteria decision support systems, it represents the acceptability of a decision after a holistic assessment of the alternatives.

Some scholars handle the response errors that may have crept into the data by specifying the functional structure of the variance of the unobserved effects. For instance, Swait and Adamowicz (1996) hypothesise that the variance associated with an element of the choice set is a function of the task complexity. Others (Bhat 1997; Hensher 1998) use respondents’ characteristics as proxies for the ability to comprehend the variability of response tasks. All choice data allow the estimation of both aggregative and individual models. The latter estimates can be aggregated at a second stage of analysis to obtain cluster estimates.

If the data are observed on an interval scale and ordered in an $(n \times p)$ matrix $\mathbf{X} = \{x_{hi} (h = 1, \dots, n; i = 1, \dots, p)\}$ with columns that sum to zero over the n obtained responses [$E(\mathbf{X}) = \mathbf{0}$], the variance-covariance matrix $\mathbf{S}$ of the items is:

² Readers may refer to Kish (1965) for a comprehensive manual.

³ The family of intra-class correlated errors also includes measures of the so-called “normative” scores (Cattell 1944), which are the scores of an individual that depend upon the scores of other individuals in the population.

$$\mathbf{S} = \mathbf{X}^T \mathbf{X} / n \quad (3)$$

where $\mathbf{X}^T$ denotes the transpose of $\mathbf{X}$. If item variances equal 1, a correlation matrix $\mathbf{R}$ is obtained. Both $\mathbf{S}$ and $\mathbf{R}$ matrices are square ($n \times n$), symmetric ($s_{ij} = s_{ji}$; $r_{ij} = r_{ji}$, $i \neq j = 1, \dots, p$) and positive semidefinite, that is $\rho(\mathbf{X}) = \rho(\mathbf{S}) = \rho(\mathbf{R})$, where ρ ($\rho \leq p$) denotes the matrix rank. Ideally, an $\mathbf{X}$ matrix, or a convenient transformation of it, is one-dimensional, so that the underlying preferences can be located along the real axis represented by the first unique factor.

The one-dimensionality of matrix $\mathbf{R}$ should be tested before the estimation of the latent values. An advance processing of the data via a principal component analysis may be used to explain the items' variance. In general, about 70 % of the observed variance should be explained by the first principal component. If the pattern exhibits more than one dimension, it may be appropriate to restrain the choice set. If an advance testing is not possible, theory on the subject matter can help the researcher in selecting the items.

If the data are observed with a scale that is at least ordinal, a preference, or dominance, matrix $\mathbf{P}$ can be computed whose diagonal elements are null and whose off diagonal cells contain measures of preference/dominance. The generic element p_{ij} ($i \neq j = 1, \dots, p$) of $\mathbf{P}$ is a measure of the dominance of the row item i over the column item j .

The properties of $\mathbf{P}$ depend on the data-collection technique. If p_{ij} assumes values constrained between 0 and 1, the symmetric element is its complement to one ($p_{ji} = 1 - p_{ij}$). If $p_{ij} = 0.5$, no preference can be expressed between i and j . If p_{ij} is larger than 0.5, it means that $i > j$ (i dominates j), and the closer the value to 1, the larger the dominance. Of course, if p_{ij} is lower than 0.5, the opposite is true; that is, $i < j$. The $\mathbf{P}$ matrix is anti-symmetric ($p_{ji} = 1 - p_{ij}$) and of full rank [$\rho(\mathbf{P}) = p$]. The estimates of ξ_i rely on the Euclidean distances between preferences (Brunk 1960; Brauer and Gentry 1968).

If Saaty's (1977) approach is adopted, the generic element of $\mathbf{P} = \{p_{ij}, i \neq j = 1, \dots, p\}$ may assume any positive value. The Author suggests that p_{ij} values vary between $1/k$ and k , where k is the maximum level of dominance of i over j ; for instance, if $k = 9$ and $p_{ij} = 9$ it means that i is nine times more preferable than j . The equilibrium value between i and j is $p_{ij} = 1$. The value in position (j, i) , symmetric to (i, j) , is its inverse ($p_{ji} = 1/p_{ij}$). The estimates of ξ_i rely on the ratio structure of $\mathbf{P}$.

It is possible to construct a matrix $\mathbf{P}$ with any data-collection technique presented in Sect. 2. The ordinal relationship between any two items may be transformed into a dominance relationship ($i > j$):

$$y_{hij} = 1 \quad \text{if } x_{hi} > x_{hj} \quad (h = 1, \dots, n; i \neq j = 1, \dots, p) \quad (4)$$

and 0 otherwise. A tie ($x_{hi} = x_{hj}$) implies no dominance relationship between i and j . For instance, if a pick-the-best technique were applied, the transformation of a choice into a 0/1 value would be based on the dominance relationships between the

best item (1) and all other items (i.e. $y_{h(1)j} = 1$ for all j s) or the dominance between the worst item (p) and the intermediate ones (i.e. $y_{hj(p)} = 1$ for all j s).

If a fixed-budget procedure were adopted and the researcher did not rely on the interval property of the data, the individual data would be considered ranks, and the conversion formula 4 could be applied. Besides, ratings can be downgraded to ranks by merely considering the dominance relations between any two items i and j .

We can estimate a vector of ξ_i values if $\mathbf{P}$ is irreducible; that is, if all values of $\mathbf{P}$ except those on the main diagonal are positive. According to the Perron-Frobenius theorem (Brauer 1957; Brauer and Gentry 1968), the right eigenvector $\mathbf{w}$ associated with the largest positive eigenvalue, λ_{max} , of matrix $\mathbf{P}$ is obtained as:

$$\mathbf{P}\mathbf{w} = \lambda_{max}\mathbf{w} \quad (\mathbf{w}^T\mathbf{w} = 1). \quad (5)$$

The elements of $\mathbf{w} = \{w_i, i = 1, \dots, p\}$ are proportional to those of the vector of estimates $\xi' = \{\xi_i, i = 1, \dots, p\}$. If the ξ_i s add up to one ($\sum_i \xi_i = 1$), it is possible to estimate a latent value as (Fabbris 2010):

$$\hat{\xi}_i = w_i / \sum_i w_i \quad (i = 1, \dots, p). \quad (6)$$

If the sample of respondents shows a consistent pattern, then all rows of the preference matrix are linearly dependent. Thus, the latent preferences are consistently estimated by values of a single row of the matrix but a multiplicative constant. All data-collection techniques introduced in Sect. 1 give about the same results in terms of relative importance of the choice set (Alwin and Krosnick 1985; Kamakura and Mazzon 1991). A way to discriminate between techniques is by analysing the reliability and validity of the obtainable responses. A data-collection technique is *reliable* if it produces accurate data in any circumstance. Reliability can be assessed through multiple measurements of the phenomenon and evaluation of the consistency of measures. A battery of items is *valid* if the obtainable data represent the phenomenon the researcher had in mind. The validity refers to the items and scales' construction techniques apt to represent the content domain. In the following, our considerations will be referred not to the single answers but to the overall correlation pattern and to the preference of each item over the others.

3.1 Ranking

The relationship between the items' rankings and the underlying values is:

$$P(x_i \geq x_j) = P(\xi_i \geq \xi_j) \quad (i \neq j = 1, \dots, p) \quad (7)$$

where $P(a)$ denotes the probability of the argument a . The error terms ε_i and ε_j ($i \neq j = 1, \dots, p$) correlate with each other because of the fixed total, T , of the ranks elicited from any individual [$T = p(p + 1)/2$]. This condition—called “ipsative” (from the Latin *ipse*; see Cattell 1944; Clemans 1966) or “compensative”, since a large assignment to an item is at the expense of the other items—causes the error terms to be unduly negatively correlated both within an individual (Chan and Bentler 1993) and through the whole sample (Vanleeuwen and Mandabach 2002). The ipsative condition induces into the correlation matrix $\mathbf{R}$ several undue negative correlations. These negative correlations depend on the data-collection technique and not on the effective values, ξ_i s, of the items. The factor configuration of ipsative measures may differ from the proper underlying configuration. Clemans (1966) showed that not only the factors of an ipsatised matrix $\mathbf{R}^*$

$$\begin{aligned}\mathbf{R}^* &= \mathbf{D}^{-1/2} \mathbf{S}^* \mathbf{D}^{-1/2}, \\ \mathbf{D} &= \text{diag} \{s_{ii}^* \ (i = 1, \dots, p)\}, \\ \mathbf{S}^* &= (\mathbf{I} - \mathbf{1}\mathbf{1}^T/p) \mathbf{S} (\mathbf{I} - \mathbf{1}\mathbf{1}^T/p),\end{aligned}$$

where $\mathbf{S}$ is the observed variance-covariance matrix, may differ from those of the raw data matrix $\mathbf{R}$ because of the large number of negative correlations, but also the rank of $\mathbf{R}^*$ is one less than that of $\mathbf{R}$. Moreover, a factor accounting for the methodological choices may stem from factor analysis. This factor, which could come out after two or three proper factors, is to be removed from the analysis of factors. Besides, the factor analysis of rankings highlights a number of factors lower than the number obtainable with ratings. Munson and McIntyre (1979) found that the correlations implied by the latter technique require about three times as many factors as the former technique.

A least-square solution for predicting a criterion variable with ipsative data is identical to the least-square solution where one variable of the ipsative set is removed before the regression analysis. The removal of a variable is a trick for coefficient estimation because of the singularity of ipsatised matrix $\mathbf{X}^*$. The proportion of deviance explained by the ipsatised data is measured by the squared multiple correlation coefficient,

$$R^{*2} = R^2 - \left(\sum_i^p \beta_i \right)^2 / \sum_{ij}^p \rho_{ij}^-, \quad (8)$$

where ρ_{ij}^- is the generic element of $\mathbf{R}^{-1}$, and $R^{*2} \leq R^2$. The operational difficulties caused by ipsatised data make it clear that a researcher should strive to obtain absolute, instead of ipsative or ipsatised measures.

The larger the number of items to be ranked, the less reliable are the obtainable rankings. It is easy to guess that the rankings' reliability is high for the most liked and disliked items and lower for those in the middle. It is not possible, however, to evaluate the reliability and validity of the ranks of the items that would never be chosen in any foreseeable circumstances and of those that are obscure or meaningless to respondents.

A good property of the ranking technique is that the measurement scale does not need to be anchored at the extremes. This favours this technique if the extremes are undefined; for instance, if the extremes are to be translated in a multilingual research study. If an individual's ranking is fully expressed, a "strong order" of the items is evident that may be used for either scoring or ranking the items (Coombs 1976). The ipsative property is irrelevant for the estimation of ξ_i s if they are estimated via a preference analysis.⁴

3.2 *Pick the Best/Worst*

The relationship between the items' choices and the underlying values is:

$$P(x_{(1)} \geq x_{(i)}) = P(\xi_{(1)} \geq \xi_{(i)}) \quad (i \neq (1), \dots, p) \quad (9)$$

where $x_{(1)}$ denotes the item selected as best. An analogous formula applies for $x_{(p)}$, the worst item. This procedure generates data that are "weakly ordered", since a complete ordering cannot be determined for the items (Louviere et al. 2000). Nevertheless, a preference matrix $\mathbf{P}$ can be estimated if a sufficiently large number of individuals are surveyed on a common choice set.

If the ξ_i s are non-compensatory,⁵ the pick-the-best strategy is preferable in spite of its simplicity. In all bounded-rationality contexts and, in particular, in social and psychological studies, this strategy showed particular plausibility (Martignon and Hoffrage 2002).

3.3 *Fixed Budget Partitioning*

The relationship between the preferences expressed for the items and their underlying values is:

$$E(x_i) = \xi_i + \varepsilon_i \quad (i = 1, \dots, p) \quad (10)$$

where $E(\varepsilon_i) = 0$ and $E(\varepsilon_i \varepsilon_j) \neq 0$. Since this technique is fixed total (i.e. ipsative), the response errors correlate with each other, in the negative direction. The maximum likelihood expectation of x_i is the arithmetic mean of the observed values over the

⁴ Even if an ipsative data-collection technique forces undue correlations on response errors, it may be worthwhile to "ipsatise" a data matrix $\mathbf{X}$ if the researcher is interested in the analysis of the differences between rows (i.e. between single individual's scores). A matrix may be ipsatised by adding a suitable constant to the scores of a respondent (i.e. to the scores of a row) so that all the new column scores sum to the same constant (Cunningham et al. 1977). Columns of matrix $\mathbf{X}$ may be rescaled to zero-mean and unit-variance, either before or after ipsatisation.

⁵ Suppose the ξ_i s are ordered from smallest to largest; they are non-compensatory if each ξ_i ($i > j = 1, \dots, p - 1$) is larger than the sum of all ξ_j s that are smaller than it (i.e.: $\xi_j > \sum_{i>j} \xi_i$).

respondents, irrespective of eventual ties. Although quantitative, the data collected with this technique ignore the absolute relevance of the evaluated items for a respondent. Let us consider, for instance, two respondents, A and B, whose effective levels of importance over four items are the following:

A: Family = 90, Work = 80, Hobbies = 70, Religion = 60;

B: Family = 70, Work = 60, Hobbies = 50, Religion = 40.

It is evident that A places systematically larger importance on any specific item than B. However, these differences do not transpire from the data collection if an ipsative data-collection technique is adopted. In fact, the following considerations apply:

- Their rankings are the same, the orders being independent of the intensity of the underlying construct.
- Since the distances between the items are the same, if budget constraints are introduced, the item distribution is the same for the two individuals.
- A and B differentiate the items in the same way, unless a ratio-based technique is taken instead of a difference-based one.⁶ Even the item variances and the inter-item covariance are the same for the two individuals.

3.4 Rating

The relationship between an item score and its underlying value is:

$$E(x_i) = \xi_i + \varepsilon_i \quad (i = 1, \dots, p) \quad (11)$$

where $E(\varepsilon_i) = 0$. The maximum likelihood expectation of x_i is the average of the observed values over all respondents. The hypothesis of a null correlation between error terms ($E(\varepsilon_i \varepsilon_j) = 0$) holds if the individual's responses given on previous items have no necessary effect on the following ones. This is sometimes called "procedural independence" (Louviere et al. 2000). The hypothesis of procedural independence does not hold if the order of administration affects the responses given after the first one. It may be argued, in fact, that the respondent focuses the underlying concept along the sequence of his or her given responses so that the preferences expressed from the second item on are anchored to the preference assigned to the previous items. In this case, the validity of responses improves from the first to the last response, but the reliability of responses diminishes because response errors correlate with each other.

One of the main drawbacks of the rating technique is the non-differentiation of responses. This means that respondents tend to dump responses towards an extreme

⁶ It may be hypothesised that responses possess the ratio property, which means that the ratio between two values is logically justified. The ipsative constraint implies, instead, that only the interval scale properties apply to fixed-total data.

of the scale, usually its maximum, making it difficult the discrimination between item scores (Alwin and Krosnick 1985; Krosnick and Alwin 1988; McCarty and Shrum 1997; Jacoby 2011). Other respondents can use just a few points of the offered scale, and this makes the scores dependent on respondents' styles (Harzing et al. 2009).

The peculiarity of the use of scale points by a respondent can be measured with the variance between the category-values he or she uses to rate the items, that is:

$$V_h = \sum_i^p (x_{hi} - \bar{x}_h)^2 / p \quad (h = 1, \dots, n), \quad (12)$$

where $\bar{x}_h = \sum_i^p x_{hi} / p$. This variance can be standardised with its maximum in the hypothesis that each item is assigned a different score and $p < k$, where k is the number of category-values of the scale:

$$\tilde{V}_h = V_h / \max(V_h) = 12V_h / (k^2 - 1). \quad (13)$$

The standardised score, $\tilde{V}_h$, varies between zero and one and is close to zero if just a few category-values are used by individual h to rate the choice set. The lower this variance the more the scale is prone to response style.⁷ This type of individual clustering of responses induces a positive correlation among the stated preferences. A combined method of picking the best and/or the worst items and then rating all the p items, as suggested by Krosnick and Alwin (1988), may be a solution for non-differentiation. The non-differentiation problem may be due to the low concentration of respondents while they answer the questionnaires; hence, the more demanding the questions, the larger the likelihood they will concentrate before answering. Nevertheless, several authors state that rating is superior in validity, certainly in comparison with ranking (Rankin and Grube 1980; Maio et al. 1996).

3.5 Paired Comparisons

The relationship between the frequency with which item i is preferred to item j and its underlying value, in an additive data-collection model, is the following:

$$p_{ij} = \xi_i - \xi_j + \varepsilon_{ij} \quad (i \neq j = 1, \dots, p), \quad (14)$$

where p_{ij} is the observed frequency of the choices expressed by all sample units to whom the pair (i, j) was administered:

⁷ Krosnick and Alwin (1988) propose Gini's measure if the researcher assumes a non-interval scale $V_h = 1 - \sum_j^k p_{hj}^2$ where p_{hj} is the frequency of point j in a scale of k points.

$$p_{ij} = \sum_h^n y_{hij} / n_{ij} \quad (i \neq j = 1, \dots, p), \quad (15)$$

and ε_{ij} is the error term associated with the (i, j) comparison. Basically, the response error depends on the order of administration of the pair. The response error increases with the number of questions because of the respondent's fatigue; on the contrary, as the pair administration proceeds, the respondent focuses the content more and more clearly, and his or her responses become more conscious. An order effect on responses is unavoidable regardless of the sequence of administered pairs (Brunk 1960; Chrzan 1994; *contra*: Scheffé 1952). This effect is practically impossible to detect in a single-shot survey unless the administration order is randomised. Randomisation is possible if the survey is computer-assisted. If the administration of pairs is random, the order effect on responses vanishes on the average over the sample of respondents.

A desired property of the expressed preferences is that respondents' choices be transitive across the full set of choices (Coombs 1976). This property is based on the consistency requirement between all preferences that involve the i th item for the estimation of ξ_i values. The transition rule considers the paired relations of preferences between any three items, i , j and k . It can be said that $i > j$ if $i > k$ and $k > j$, where k is any item of the choice set but i and j . In terms of matrix $\mathbf{P}$ elements:

$$p_{ij} \geq p_{ik} \quad \text{if } p_{ik} \geq p_{kj} \geq 0.5 \quad (k \neq i \neq j = 1, \dots, p), \quad (16)$$

$$p_{ij} \leq p_{ik} \quad \text{if } p_{ik} \leq p_{kj} \leq 0.5 \quad (k \neq i \neq j = 1, \dots, p). \quad (17)$$

The application of transition rules makes the tournament technique viable. The application of a transition rule to estimate a preference between two missing pairs implies the computation of the estimates of p_{ij} through all possible k s and the synthesis (usually the average) of the estimated preferences. Transition rules may also be used to evaluate the internal consistency of a $\mathbf{P}$ matrix. It is possible, in fact, that any three preferences p_{ij} , p_{ik} and p_{kj} are not mutually coherent. Any inconsistency among such a triad may be considered a symptom either of a measurement error or of (logical) inner incoherence. The inconsistency of a set of preferences can be hypothesised as proportional to the number of triads that do not satisfy the rank-order preservation rule. Fabbri (2011) proposes an index of inconsistency based on the ratio between the number of incoherent triads and its maximum value under the full inconsistency hypothesis.

4 Practical Considerations

The feasibility of the preference elicitation techniques will be examined from the following viewpoints:

- *Respondent burden*; that is, time and physical and psychological effort required from responding individuals.
- *Expected proportion* of missing values. It is a measure of the items' expected difficulty or of negative attitudes toward the survey content among the interviewees. The higher the proportion of missing values, the higher the risk that the collected data ignore specific segments of the surveyed population.
- *Possibility of application* with a face-to-face (CAPI) or telephone (CATI) interviewing system, or any kind of online procedure (e.g. through the web, email or other online procedures (CAWI), self-administered (CASI) or mail questionnaire).

The same order of presentation of the preference elicitation techniques as Sect. 2 will be followed.

4.1 Ranking

The difficulty of the respondent task increases substantially with p , the number of items. In fact, it can be guessed that a typical respondent reads and compares all the items to identify the most and/or the least important items. Then, a series of multiple comparisons is required for respondents to define the second best and the following positions. The “best-worst” procedure introduced in Sect. 2.1 may be a rationalisation of the foregoing dynamics. Even if it is generally unknown how the respondent performs his or her task, the best-worst sequence can reduce respondent acquiescence (McCarty and Shrum 1997).

What makes the ranking technique generally acceptable is that respondents are accustomed to ranking provided ties are allowed and the number of items is reasonable. Several scholars (including Rokeach and Ball Rokeach 1989, among others), balancing the likely response burden with the wealth of information contained in a ranking, and assuming that this technique forces participants to differentiate between similarly regarded techniques, propose ranking as the preferable technique regardless of the self-administered survey mode (CAWI, CASI, mail).

Ranking is not feasible in a CATI survey, because it is necessary for the respondent to examine repeatedly the set of alternatives, and this is feasible only if the respondent can memorise the item list. Whatever the device applied to exhibit the item list, the results differ according to respondent personality; also, embarrassing pauses and unwanted hints from the interviewer may condition the respondent's task. Even in a face-to-face survey, the presence of an interviewer may bother or embarrass the respondent during such a task. This difficulty may be attenuated if interviewers are specifically trained.

If respondents are insufficiently motivated to collaborate, the extreme choices (e.g. the top and bottom items) may be accurately elicited, but the intermediate positions may be uncertain. That is why under a difficult survey conditions (for instance, if the data are collected in a hurry or the interviewee is standing), the simple pick-the-best/worst technique is advisable. Another difficult task may be that of collecting data

on non-choices (e.g. the ordering of a set of alternatives the respondent would never choose). Louviere et al. (2000) suggest not considering ranking as an appropriate procedure in this regard.

4.2 Picking the Best/Worst

This procedure is the least-burdening among those that require respondents to evaluate a whole set of alternatives at a time, for just the extremes are to be chosen. It may be considered a truncated form of the complex tasks required by the ranking technique. It can be applied easily with any interviewer-based (CAPI) or self-administered (CASI, CAWI, mail) questionnaire. It can be applied in a CATI survey only if the list to be remembered is short.

4.3 Fixed Total Partitioning

The fixed total technique is burdensome and time consuming for respondents, because it requires not only the ranking of the item set but also the quantitative valuing of each item. If the choice set is large, it is possible to break it into categories and then ask respondents to prioritise the items within each category.

The responses obtained with the fixed total procedure are subject to response style, since it remains unknown how a respondent performs the required multiple task; namely, ranking the alternatives, assigning them a value, checking the score total and adjusting the scores according to the given budget. As an example, it may happen that a mathematically minded respondent starts his or her task by dividing the budget by the number of items to have a first idea of the average score and then picks the top item and possibly the second best or the bottom item with a mechanism similar to that of the picking-the-best technique. Then, he or she has to rate them, maybe by choosing a differentiation-from-average criterion. It is clear that the task is highly burdensome, more than any ranking or rating procedure, and that the intermediate scores are precarious.

Hence, this technique is prohibitive in CATI surveys for the same reasons put forward for ranking. It is plausible that respondents collaborate in full if an interviewer is present where the respondents gather (CAPI mode) and are motivated to collaborate in the survey.

It is an acceptable procedure for self-administered surveys provided the respondent population is sympathetic to the research aims and the questionnaire is communicative. In self-administered surveys, in fact, the respondents try to seek information from the instrument itself (Ganassali 2008).

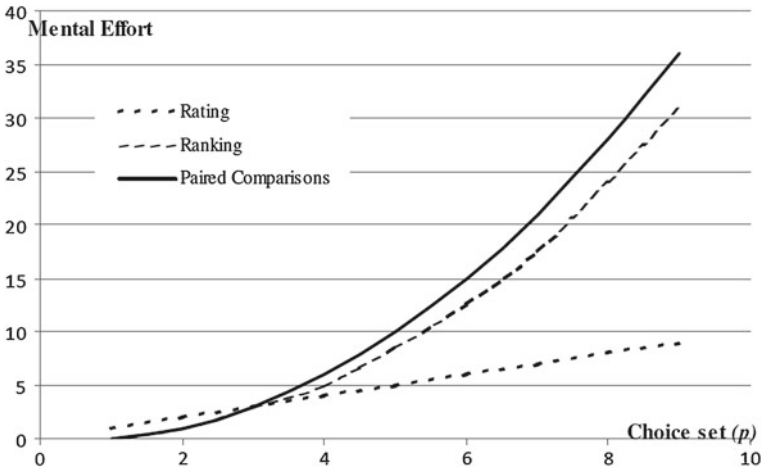


Fig. 7 Respondents' mental effort hypothesised in performing ranking, rating and paired comparison procedures

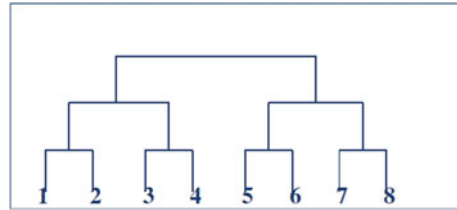
4.4 Rating

Rating single items is the easiest procedure among those presented in Sect. 2, since people are asked to assign a value to each alternative and the time and fatigue associated with their task is about linear (i.e. proportional to the number of items). The superiority of the rating versus the ranking procedure from the viability viewpoint becomes larger and larger as the number of items diverges. In Fig. 7, the lower straight line represents the expected respondent fatigue if the rating procedure is applied, and the dotted curve represents⁸ that of ranking.

The two techniques diverge notably as burden is concerned for p larger than three or four. The curve representing the burden of the fixed-total procedure might be even steeper than ranking. However, ratings are prone both to social desirability biases and to difficult discrimination between items. For instance, suppose that customer preferences for innovations in urban bus transportation are to be elicited and juxtapose the alternative of modifying the bus structure to favour the boarding of the disabled and that of increasing the night services for all customers. It is reasonable to expect high scores for both items (devices for the disabled and night services) if rating is applied; however, night services might get much higher preferences if either ranking or pair comparison is applied. In fact, a respondent's choice that refers to his or her

⁸ The curve representing the effort due to ranking was traced by assuming that reading, memorising and ordering an item with respect to the previous item (ranking task) implies to a respondent the same fatigue as reading and assigning a score to an item (rating task). The ideal respondent should repeat the set of memorisation and ordering for all items but those already ranked. The burden guessed for the ranking task is probably an underestimate of the fatigue a conscious respondent would put in performing his or her task. Munson and McIntyre (1979) state that ranking takes about three times longer than comparable rating.

Fig. 8 Hierarchical tournament structure, $p = 8$



own utility tends to discriminate among the possible items, while the independent evaluation of a single item may be conditioned by social desirability. Moshkovich et al. (1998) even found that if items are important to respondents, they tend to rate them lower than if items are insignificant to them, since they hypothesize that people tend to score highly if the subject matter is socially marginal.

4.5 Paired Comparisons

The choice of the preferred alternative when items are administered in pairs is very burdensome. It requires a lot of respondents' time and causes mental fatigue, because a choice is to be made for all distinct pairs. The difficulty associated with the respondent's task is larger if the paired items logically overlap or if their marginal rate of substitution is high (Bettman et al. 1993). This implies that this procedure cannot be applied in full if p is larger than four, unless respondents are faithful. In Fig. 7, it is hypothesized that paired comparison is more burdensome than ranking.

If p is large, it is possible to classify the items in homogeneous categories and perform first the paired comparisons within each category and then compare the categories. This makes the direct comparison between the unmatched items impossible.

Another reduced version of paired comparisons, such as the tournament technique or an incomplete experimental design, is usually necessary regardless of the survey mode. Figure 8 represents an administration scheme of the latter technique with $p = 8$. More complex structures can be applied for a number p of items that is a multiple of either two or three (Fabbris and Fabris 2003). The tournament technique requires the administration of just $(p - 1)$ questions, instead of $p(p - 1)/2$, and is acceptable if the initial coupling of items is randomised. The tournament technique may be set up for any kind of computer assisted interviewing if an expert is available for computer programming. Programming assistance is required, in particular, because the randomisation of pairs implies the necessity of typing the unordered responses within appropriate fields in the database. The random presentation of pairs to the respondent is also a solution for eliminating the order effect if the complete paired comparison procedure is adopted. Hence, a computer-assisted data-collection system is particularly appropriate regardless of the choice-based technique, adminis-

tered either with the mediation of an interviewer or without it. The paired comparison technique is not viable in remote surveys (CATI, CAWI, mail).

5 Concluding remarks

The most diffused techniques for collecting data on sets of interrelated items have been examined. The analysis was held with reference to both statistical and practical paradigms. No technique showed enough positive properties to make it preferable in all survey situations. The feasibility of the scales was based chiefly on the burden posed on respondents to perform their task. The respondent burden correlates with the quantity, reliability and validity of the information expected from responses; in particular, the more demanding the task, the lower the sample coverage and the higher the frequency and variety of errors. Therefore, the researcher should make his or her methodological choices balancing the tolerable burden and the expected information according to the population at hand.

Choosing the appropriate scale requires an analysis of the research context at the data-collection design stage. The analysis of people's propensity to collaborate during data collection is a crucial issue for the application of burdensome techniques. The choice of technique will definitely affect the quality of the collected data and hence the survey outcomes.

Moreover, each technique, or a convenient transformation of it, was associated to a computer-assisted data-collection mode. The more complex techniques (i.e. those based on ranking, pair-comparison of the items or on budget partitioning among the items) are inappropriate for CATI surveys but may be applied in self-administered data-collection modes (CASI, CAWI). In a face-to-face survey, a ranking-based technique may be adopted with the appropriate training of interviewers. The presence of an interviewer may support the administration of paired comparisons or of a reduced type of paired comparisons such as the tournament technique.

The simpler techniques are that of picking the best/worst items and that of rating the items. Let us concentrate on the latter, because many authors advise using it because of its practical simplicity and the validity of the obtainable estimates. If the mechanics of the duties implied by the ranking and rating procedures are compared, it can be concluded that they realise two different aims. Ranking applies to studies whose general aim is to resolve conflicts among the respondents' interests, classifying items and taking decisions, since the inner pattern is measured by the comparative scale of the concerned items. The rating procedure, conversely, enables researchers to elicit the respondents' structural values in an unconstrained manner (Ovada 2004). This difference also applies if paired comparisons and the rating technique are compared (Aloysius et al. 2006). Hence, a rating does not imply a contrast between the items but the "absolute" importance of each item with respect to the subject matter. The rating procedure is inadequate for item classification and decision making, because it is exposed both to social desirability bias and to non-discrimination between the items. The responses obtainable with different techniques

may be conditioned by the respondents' culture as far as culture correlates with the task difficulty and responding availability. If the researcher suspects that either the participation level or the quality of responses may depend on people's culture, easier tasks (namely, the rating technique, short lists of alternatives or a qualitative rather than a quantitative scale) should be applied. Besides, more research is needed to control the effects of culture-conditioned errors on responses.

Let us consider the case in which ratings are transformed into paired comparisons by considering the non-null contrasts between any two items. An analysis of the dominance matrix constructed with the informative contrasts may give more discriminatory estimates of the items' effective values than those obtainable by processing the observed ratings directly. This trick could overcome the non-differentiation problem, but it is unable to eliminate the social desirability effect (see also Chapman and Staelin 1982; Louviere et al. 1993).

Even if a technique satisfies the researcher's aims and the context conditions, he or she should consider collecting, and then processing, a combination of data obtained with two or more techniques. A combination of multiple-source data is sometimes called "data enrichment" (Ben-Akiva et al. 1994), because any data-collection technique captures aspects of the choice process for which it is more appropriate.⁹ The adoption of either a ranking-based or a picking-the-best/worst technique, before a rating, may induce the respondents to adopt comparative response logics that they generalise to the subsequent rating task, and this may attenuate the effect of anchoring responses to the scale extremes while exposing the respondents to the full range of the possible values before rating the items. McCarty and Shrum (2000) showed that the application of the simple most-least procedure before a rating took one-third more time than usually required by a simple rating procedure.

Research on preference regularity is under way. Regular is a methodological outcome that is common to two or more preference elicitation techniques. Louviere et al. (2000) applied the regularity criterion to the parameters they estimated and stated that two methods are regular if they exhibit parameters that are equal up to a positive constant. They found that parameter estimates are regular in a surprisingly large number of cases.

In conclusion, a meta-analysis of empirical studies and experiments on real populations would be welcome to shed light on the relationships between the statistical properties and the practical and purpose-related feasibility of techniques. In fact, the results of experiments vary according to contents and contexts. Thus content- and context-specific rules should be determined that are appropriate to survey peculiar segments of the target population. Instead, the experiments conducted on convenience samples, such as on-campus students, which may inform on estimates' validity are irrelevant for both reliability and feasibility judgements on survey techniques.

⁹ Data-collection techniques alternative to the basic ones presented in Sect. 2 are rare but increasing in number. Louviere et al. (2000) quote experiments by Meyer (1977), Johnson (1989), Louviere (1993) and Chrzan (1994).

Acknowledgments This paper was realised thanks to a grant from the Italian Ministry of Education, University and Research (PRIN 2007, CUP C91J11002460001) and another grant from the University of Padua (Ateneo 2008, CUP CPDA081538).

References

- Aloysius, J. A., Davis, F. D., Wilson, D. D., Taylor, A. R., & Kottemann, J. E. (2006). User acceptance of multi-criteria decision support systems: The impact of preference elicitation techniques. *European Journal of Operational Research*, 169(1), 273–285.
- Alwin, D. F., & Krosnick, J. A. (1985). The measurement of values in surveys: A comparison of ratings and rankings. *Public Opinion Quarterly*, 49(4), 535–552.
- Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143–156.
- Bassi, F., & Fabbris, L. (1997). Estimators of nonsampling errors in interview-reinterview supervised surveys with interpenetrated assignments. In: L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 733–751). New York: Wiley.
- Beatty, S. E., Kahle, L. R., Homer, P., & Misra, K. (1985). Alternative measurement approaches to consumer values: The list of values and the Rokeach value survey. *Psychology and Marketing*, 2(3), 81–200.
- Ben-Akiva, M. E., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppenwal, H., & Rao, V. (1994). Combining revealed and stated preferences data. *Marketing Letters*, 5(4), 335–351.
- Bettman, J. R., Johnson, E. J., Luce, M. F., & Payne, J. (1993). Correlation, conflict, and choice. *Journal of Experimental Psychology*, 19, 931–951.
- Bhat, C. (1997). An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science*, 31(1), 34–48.
- Brauer, A. (1957). A new proof of theorems of Perron and Frobenius on nonnegative matrices. *Duke Mathematical Journal*, 24, 367–368.
- Brauer, A., & Gentry, I. C. (1968). On the characteristic roots of tournament matrices. *Bulletin of the American Mathematical Society*, 74(6), 1133–1135.
- Brunk, H. D. (1960). Mathematical models for ranking from paired comparisons. *Journal of the American Statistical Association*, 55, 503–520.
- Cattell, R. B. (1944). Psychological measurement: Normative, ipsative, interactive. *Psychological Review*, 51, 292–303.
- Chan, W., & Bentler, P. M. (1993). The covariance structure analysis of ipsative data. *Sociological Methods Research*, 22(2), 214–247.
- Chapman, R., & Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, 19(3), 288–301.
- Chrzan, K. (1994). Three kinds of order effects in choice-based conjoint analysis. *Marketing Letters*, 5(2), 165–172.
- Clemans, W. V. (1966). *An analytical and empirical examination of some properties of ipsative measures*. *Psychometric monographs* (Vol. 14). Richmond: Psychometric Society. <http://www.psychometrika.org/journal/online/MN14.pdf>.
- Coombs, C. H. (1976). *A theory of data*. Ann Arbor: Mathesis Press.
- Conrad, F. G., Couper, M. P., Tourangeau, R., & Galesic, M. (2005). *Interactive feedback can improve the quality of responses in web surveys*. In: ESF Workshop on Internet Survey Methodology (Dubrovnik, 26–28 September 2005).
- Crosby, L. A., Bitter, M. J., & Gill, J. D. (1990). Organizational structure of values. *Journal of Business Research*, 20, 123–134.

- Cunningham, W. H., Cunningham, I. C. M., & Green, R. T. (1977). The ipsative process to reduce response set bias. *Public Opinion Quarterly*, 41, 379–384.
- Elrod, T., Louviere, J. J., & Davey, K. S. (1993). An empirical comparison of ratings-based and choice-based conjoint models. *Journal of Marketing Research*, 24(3), 368–377.
- Fabbbris, L. (2010). Dimensionality of scores obtained with a paired-comparison tournament system of questionnaire items. In: F. Palumbo, C. N. Lauro & M. J. Greenacre (Eds.), *Data analysis and classification. Proceedings of the 6th Conference of the Classification and Data Analysis Group of the Società Italiana di Statistica* (pp. 155–162). Berlin: Springer.
- Fabbbris, L. (2011). One-dimensional preference imputation through transition rules. In: B. Fichet, D. Piccolo, R. Verde & M. Vichi (Eds.), *Classification and multivariate analysis for complex data structures* (pp. 245–252). Heidelberg: Springer.
- Fabbbris, L., & Fabris, G. (2003). Sistema di quesiti a torneo per rilevare l'importanza di fattori di customer satisfaction mediante un sistema CATI. In: L. Fabbbris (Ed.), *LAID-OUT: Scoprire i Rischi Con l'analisi di Segmentazione* (p. 322). Padova: Cleup.
- Fellegi, I. (1964). Response variance and its estimation. *Journal of the American Statistical Association*, 59, 1016–1041.
- Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. *Survey Research Methods*, 2(1), 21–32.
- Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54(4), 3–19.
- Green, P. E., Krieger, A. M., & Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, 31(3.2), 556–573.
- Gustafsson, A., Herrmann, A., & Huber, F. (Eds.). (2007). *Conjoint measurement: Methods and applications* (4th edn.). Berlin: Springer.
- Harzing, A.-W., et al. (2009). Rating versus ranking: What is the best way to reduce response and language bias in cross-national research? *International Business Review*, 18(4), 417–432.
- Hensher, D. A. (1998). Establishing a fare elasticity regime for urban passenger transport: Non-concession commuters. *Journal of Transport Economics and Policy*, 32(2), 221–246.
- Jacoby, W. G. (2011). *Measuring value choices: Are rank orders valid indicators?* Presented at the 2011 Annual Meetings of the Midwest Political Science Association, Chicago, IL.
- Johnson, R. (1989). Making decisions with incomplete information: The first complete test of the inference model. *Advances in Consumer Research*, 16, 522–528.
- Kamakura, W. A., & Mazzon, J. A. (1991). Value segmentation: A model for the measurement of values and value systems. *Journal of Consumer Research*, 18, 208–218.
- Kish, L. (1965). *Survey sampling*. New York: Wiley.
- Krosnick, J. A., & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. *Public Opinion Quarterly*, 52(4), 526–538.
- Kuhfeld, W. F., Tobias, R. B., & Garratt, M. (1994). Efficient experimental design with marketing research applications. *Journal of Marketing Research*, 31(4), 545–557.
- Louviere, J. J., Fox, M., & Moore, W. (1993). Cross-task validity comparisons of stated preference choice models. *Marketing Letters*, 4(3), 205–213.
- Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: Analysis and application*. Cambridge: Cambridge University Press.
- Maio, G. R., Roese, N. J., Seligman, C., & Katz, A. (1996). Rankings, ratings, and the measurement of values: Evidence for the superior validity of ratings. *Basic and Applied Social Psychology*, 18(2), 171–181.
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparisons. *Theory and Decisions*, 52, 29–71.
- McCarty, J. A., & Shrum, L. J. (1997). Measuring the importance of positive constructs: A test of alternative rating procedures. *Marketing Letters*, 8(2), 239–250.
- McCarty, J. A., & Shrum, L. J. (2000). The measurement of personal values in research. *Public Opinion Quarterly*, 64, 271–298.

- Meyer, R. (1977). An experimental analysis of student apartment selection decisions under uncertainty. *Great Plains-Rocky Mountains Geographical Journal*, 6(special issue), 30–38.
- Moshkovich, H. M., Schellenberger, R. E., & Olson, D. L. (1998). Data influences the result more than preferences: Some lessons from implementation of multiattribute techniques in a real decision task. *Decision Support Systems*, 22, 73–84.
- Munson, J. M., & McIntyre, S. H. (1979). Developing practical procedures for the measurement of personal values in cross-cultural marketing. *Journal of Marketing Research*, 16(26), 55–60.
- Ovada, S. (2004). Ratings and rankings: Reconsidering the structure of values and their measurement. *International Journal of Social Research Methodology*, 7(5), 404–414.
- Rankin, W. L., & Grube, J. W. (1980). A comparison of ranking and rating procedures for value system measurement. *European Journal of Social Psychology*, 10(3), 233–246.
- Rokeach, M. (1967). *Value survey*. Sunnyvale: Halgren Tests (873 Persimmon Avenue).
- Rokeach, M. (1979). *Understanding human values: Individual and societal*. New York: Free Press.
- Rokeach, M., & Ball Rokeach, S. J. (1989). Stability and change in American value priorities. *American Psychologist*, 44, 775–784.
- Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15, 234–281.
- Scheffé, H. (1952). An analysis of variance for paired comparisons. *Journal of the American Statistical Association*, 47, 381–400.
- Small, K. A., & Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, 49(1), 105–130.
- Swait, J., & Adamowicz, W. (1996). *The effect of choice environment and task demands on consumer behavior: Discriminating between contribution and confusion*. Department of Rural Economy, Staff Paper 96-09, University of Alberta, Alberta.
- Vanleeuwen, D. M., & Mandabach, K. H. (2002). A note on the reliability of ranked items. *Sociological Methods Research*, 31(1), 87–105.

Survey Data Collection and Integration

Davino, C.; Fabbris, L. (Eds.)

2013, XII, 156 p., Hardcover

ISBN: 978-3-642-21307-6