Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v8n3p329

Modelling scale effects and uncertainty in rating surveys
By Iannario

20 December 2015

Electronic Journal of Applied Statistical Analysis
Vol. 08, Issue 03, December 2015, 329-345
DOI: 10.1285/i20705948v8n3p329

Modelling scale effects and uncertainty in rating surveys

Maria Iannario*

University of Naples Federico II, Department of Political Sciences
Via L. Rodinò, 22, 80138, Naples, Italy

20 December 2015

In social and behavioural sciences the optimal use of rating scales is an important issue. Discussions on the number of response categories that optimizes the psychometric proprieties of the scales and sums up the global amount of information deriving by the response patterns are the aim of several studies. In a parametric context, this paper faces two specific topics: verify if the number of modalities affects the interpretation of responses and suggest a transformation of the observed ratings distribution when a comparison of results with a different numbers of response categories is necessary. Analysis is carried out within the statistical framework introduced for the study of feeling and uncertainty components in the process of responses which generate ordinal scores. Some empirical evidences and a simulation experiment support the usefulness of the approach.

Keywords: Ordinal response models, Ratings, Scale effects, CUB models.

1 Introduction

In several surveys the interpretation of rating scales introduced for analyzing individual evaluations, attitudes and opinions is a common and important issue. Discussions have been conducted to examine the effects of different numbers of response categories on the reliability and validity of rating scales and on the optimal number of response alternatives (Cox, 1980; Cicchetti et al., 1985; Schutz and Rucker, 1975; Matell and Jacoby, 1971; Lozano et al., 2008).

From a different point of view, other recent approaches focus on the interpretation of the scales with same number of categories. They are concentrated on subjective

*Corresponding author: maria.iannario@unina.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

interpretation of ordinal scale by considering the *anchoring vignettes*. These are (usually brief) descriptions of hypothetical people or situations that survey researchers can use to correct inter-personally incomparable survey responses (King et al., 2004; King and Wand, 2007).

Analyses also concern the structure and the choice of the scale. With respect to questions where the respondent is asked to provide the agreement to a statement (Likert, 1932) which provides a description of ordered response levels (for example: strongly disagree, disagree, neither agree nor disagree, agree, strongly agree), the scale used to measure specific ordinal issues may be also a numerical one.

In the former category, several researchers use the "Equal-Appearing Intervals" scaling method (Thurstone, 1927) to produce what is argued to be an interval scale.

In the second category, related to the choice of the scale, Cummins and Gullon (2000) state that using an expanded scale is desirable for subjective quality of measurement and that the appropriate scale format may be a 10-point, end-defined scale. Generally, several dimensions (more than 7) of scales allow to deal factors as social desirability, response set (stereotyped answers), semantic factors. Garner (1960) suggests that maximum information is obtained by using more than 20 response categories. Other Authors argue that reliability is maximized with 7 response alternatives (Symonds, 1924; Ramsay, 1973; Nunnally, 1970; McKelvie, 1978) or 5 (Jenkins and Taber, 1977; Lissitz and Green, 1975; Neumann, 1979) by stressing the idea than a grater categorization implies a greater loss of information. The general thought is that the discriminating power is lowest for the scales with a low number of categories (less than 4) (Preston and Colman, 2000) because respondents prefer formats with a larger number of response sets, as this permits them to more clearly express their point of view. Finally, Cox (1980) argues that the number 7 plus or minus 2 appears to be a reasonable range for the optimal number of response alternatives.

Some authors argued that the response may vary if different number scale points are used (Dawes, 2008) since this influences what is commonly denoted in psychological literature as *response style* (Baumgartner and Steenback, 2001; Buckley, 2009; Cronbach, 1998).

Moreover, respondents may not perceive all the adjacent levels as equidistant. Thus, when using a 10 values scale it can be argued that the psychological distance between 5 and 6, that in some countries is associated to the pass mark, can be larger than those associated to other adjacent values. Another form of distortion may arise from the fact that respondents may refrain from using extreme values of the scale when a more generous number of scale points is provided, while they are usually more concentrated in the central value. Middle category endorsement in odd-number scale is extensively analyzed by Kulas and Stachowski (2009). Mostly, odd numbers of response categories have generally been preferred to even numbers because they allow the middle category to be interpreted as a neutral point (Green and Rao, 1970; Neumann and Neumann, 1981).

An extensive literature, instead, discusses about the inclusion of *don't know* option in a response scale (see the seminal work of Converse (1964), and Beatty and Herrmann (1995); Poe et al. (1988); Lietz (2010), among others).

Another matter related to the scale effects concerns the comparison. It is sometimes found that data on the same topics are collected by means of different scales. This circumstance creates serious problems in case of comparison. In longitudinal research designs, in analysis concerning different countries or simply in studies related to different fields as education, medicine, marketing, or many areas of social research, for example, a m-point scale may be replaced by a new m^*-point scale, or vice versa, and researchers may wish to establish a basis for continuity to enable comparisons to be made among different data.

In this paper we confine our attention to the analysis and comparability of scales with more than 3 categories. Specifically, we analyze the scale usage problems in two different perspectives:

- verify if the number of categories affects the interviewed's response;
- introduce a model-based method for transforming the expressed responses on different scales.

To deal with both issues a mixture model for the analysis of ordinal variables will be introduced. The basic idea is to capture the fundamental components which are present in the process of selection of a level/category in a parsimonious manner by means of a parametric probability distribution introduced for the analysis of ordinal variables. The idea behind the mixture model is to weigh the feeling and uncertainty components present in the process of selection of a rating score. These features deserve careful consideration since they turn out to be useful in different circumstances when ordinal data have to be compared.

The paper is organized as follows: in Section 2 the setting and notation of the selected mixture model is presented. Section 3 discusses the effect of changing a scale on the responses' distribution and Section 4 performs a simulation experiment to enforce the empirical evidence. Then, Section 5 introduces a proposal to compare rating distributions obtained by different scales according to a model-based approach. Some final remarks end the work.

2 Method and setting

The finite discrete mixture which is the main tool of this paper has been denoted as CUB model (the acronym stems from the *C*ombination of a discrete *U*niform and shifted *B*inomial random variable) and it has been introduced by Piccolo (2003) to analyze ordinal data as the result of a data generating process (Iannario and Piccolo, 2012, 2015). This approach differs from the more common models for ordinal data (for instance, cumulative models: McCullagh (1980); Agresti (2010); Tutz (2010)) since it explicitly considers the discrete probability of the ordinal categories without considering the cumulative probabilities derived from the latent variable. As a consequence, CUB models are *ceteris paribus* more parsimonious since they do not require the estimation of thresholds in the fitting procedure. In addition, as we will discuss later, parameters are immediately related to the psychological components of human decisions and a visual representation simplifies this interpretation (see Gambacorta and Iannario (2013), for a comparison of the two approaches).

The starting point for introducing CUB models is a formal setting of the behaviour of respondents when faced with rating scales. Generally, the interviewee performs a comparative judgment based on his/her background, interest, personal feeling, attractiveness, satisfaction, awareness towards the item. At an unconscious level, he/she evaluates the desirability of responding accurately (the so-called "communicative intent" of Bradbum et al. (1979)), but possible indecision, fuzziness and blurriness quite often surround the final selection.

To simplify such a complex cognitive process we refer to the main factors present in the selection of an ordinal category as *feeling* and *uncertainty*, respectively, and advocate for them two discrete probability mass functions. Assume that each subject generates an ordinal response Y_i and $Pr\,(Y_i = j)$ is the probability that he/she selects the j-th category, where j belongs to the support $\{1, 2, \dots, m\}$ and m is a known integer. All information on the i-th subject are collected in the i-row $\boldsymbol{t}_i$ of a matrix $\boldsymbol{T}$, for $i = 1, 2, \dots, n$.

The *stochastic component* of a CUB model is defined by:

$$Pr(Y_i = j \mid \boldsymbol{\theta}) = \pi_i\, b_j(\xi_i) + (1 - \pi_i)\, p_j^U\,, \qquad i = 1, 2, \dots, n, \;\; j = 1, 2, \dots, m, \tag{1}$$

where $b_j\,(\xi_i) = \binom{m-1}{j-1} \xi_i^{m-j}\,(1 - \xi_i)^{j-1}$ and $p_j^U = 1/m$, $j = 1, 2, \dots, m$ are the probability mass functions of the shifted Binomial and discrete Uniform random variables, respectively.

Then, the *systematic components* of the model link parameters and subjects in a direct way by means of:

$$logit(\pi_i) = \boldsymbol{x}_i \boldsymbol{\beta}; \qquad logit(\xi_i) = \boldsymbol{w}_i \boldsymbol{\gamma}; \qquad i = 1, 2, \dots, n, \tag{2}$$

where $(\boldsymbol{x}_i, \boldsymbol{w}_i)$ is the information set extracted from the i-th row $\boldsymbol{t}_i$ of $\boldsymbol{T}$ to specify the relationships of π_i and ξ_i with the corresponding subjects' covariates $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$. The logistic function $logit(p) = \log\,(p/(1 - p))$, for any real $p \in (0, 1)$ is generally the most common link function. Notice that, given the parameterization (2), the covariates in $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ may coincide, overlap or be completely different.

To interpret the parameters of the model, let us consider that the *propensity* of the subject to give a completely random response increases with $(1 - \pi_i)$ whereas the agreement with the item increases with $(1 - \xi_i)$ since $\xi_i \to 0$ implies a distribution which gives high probabilities to greater values of the categories. As a consequences, we may interpret $(1 - \pi_i)$ and $(1 - \xi_i)$ as direct measures of *feeling* and *uncertainty* of the i-th subject, respectively.

When the model is fitted to the whole sample and without covariates, we get $\pi_i \equiv \pi$ and $\xi_i \equiv \xi$, $\forall i$ and a global model for the sample of respondents becomes:

$$Pr(Y = j \mid \boldsymbol{\theta}) = Pr_j\,(\boldsymbol{\theta}) = \pi b_j\,(\xi) + (1 - \pi)\, p_j^U, \qquad j = 1, 2, \dots, m. \tag{3}$$

Now, $\boldsymbol{\theta} = (\pi, \xi)'$ belongs to the parameter space

$$\Omega\,(\boldsymbol{\theta}) = \{(\pi, \xi) : \;\; 0 < \pi \leq 1,\, 0 \leq \xi \leq 1\}\;.$$

The identifiability of model (2) has been proved for any $m > 3$ (Iannario, 2010) whereas $m = 3$ implies a saturated model. Since the parameter space is the (left open) unit square and there is a one-to-one correspondence between the probability mass function (3) and a point in the unit square with coordinates $(1 - \pi,\, 1 - \xi)$, the interpretation of CUB model is immediate. In particular, it allows for an easy comparison when different models are built with respect to subjects' characteristics (gender, education, marital status, job, etc.) or varying time, space and contexts.

Asymptotic inference related to estimation and testing for CUB models are obtained by Maximum Likelihood methods and EM procedures for finite mixtures, as fully discussed by Piccolo (2006). An updated program to perform a full statistical inference for CUB models and several variants is freely available in the R environment (Iannario and Piccolo, 2014).

Moreover, we mention that several extensions have been advanced for CUB models as, for instance, the inclusion of objects' covariates (Piccolo, and D'Elia 2008), a *shelter* component (Iannario, 2012a), a multilevel option (Iannario, 2012b), the presence of overdispersion (Iannario, 2015) and the modelling of "don't know" responses (Manisera and Zuccolotto, 2014).

3 Changing the scales and uncertainty component

In this Section we assume that preference and agreement towards the item are collected as continuous covariates and, for practical reasons, people split their range by selecting a specific option within a discrete set of ordinal categories. It is legitimate to ask if the feeling and uncertainty measures are modified by such a subdivision: i) *a priori*, to suggest a convenient number of categories when the survey has to be planned; ii) *a posteriori*, to assess a correct interpretation of the results for a given subdivision.

If CUB models are considered as an adequate parameterization for the analysis of ordinal data, a general result may be derived. In fact, the parameter π is strictly related to the uncertainty which may be also interpreted in terms of heterogeneity of the distribution. A discrete Uniform distribution is the extreme case of a totally uncertain response and it corresponds to the maximum heterogeneity (such a random variable attains the maximum entropy among all discrete distributions defined over a finite support). On the contrary, uncertainty and heterogeneity both decrease if the distribution is concentrated in a single category or a limited number of categories.

For a discrete random variable X characterized by the probability distribution function $(p_1, p_2, \dots, p_m)$ over the support $\{1, 2, \dots, m\}$, a convenient measure of this concept is the (normalized) Gini heterogeneity defined by:

$$G_X = \left(1 - \sum_{j=1}^{m} p_i^2\right) \frac{m}{m-1}\,, \tag{4}$$

which satisfies: $0 \leq G_X \leq 1$. Then, for a CUB model, it has been shown (Iannario, 2009, 2012c) that

$$G_{CUB} = 1 - \pi^2\,(1 - G_{SB})\,, \tag{5}$$

where G_{CUB} and G_{SB} are the Gini indexes for CUB and shifted Binomial random variables, respectively. The last identity shows that, for a given ξ, heterogeneity (as measured by the Gini index) is inversely related to π and thus it is directly related to uncertainty (as measured by $1-\pi$).

Now, let X and W two random variables defined over the supports $\{1, 2, \dots, m\}$ and $\{1, 2, \dots, m, m+1\}$ with probability mass functions $(p_1, p_2, \dots, p_m)$ and $(q_1, q_2, \dots, q_m, q_{m+1})$, respectively. Assume that $p_j \equiv q_j$, $j = 1, 2, \dots, m-1$ whereas $p_m = q_m + q_{m+1}$. Thus, the two distributions are identical but for the last category of X whose probability is split over the last two categories in the case of W. Then, it is immediate to prove that:

$$G_W = G_X + 2q_m\, q_{m+1} \geq G_X .$$

This result shows that, in this instance, an increase in the number of categories increases the Gini index, therefore heterogeneity and the uncertainty of discrete distributions as measured by $1-\pi$ for CUB models.

To check the empirical validity of this result we consider two real case studies where a continuous latent variable has been codified in a finite number m of categories. Then, for varying m, a CUB model has been fitted to these ordinal data and a sequence of estimated models has been plotted in the parameter space, for $m = 3, 4, \dots, 20$.

- The first case study concerns a self-administered questionnaire collected in 2014 in which a subsample of 1108 respondents rated some issues associated to relational goods (see Gui and Stanca (2010); Uhlaner (1989), among others). Since time spent for relationships has a significant and positive impact on individual happiness, the analysis is referred to the self declared *happiness*. This variable is collected by means of a continuous line (track bar) with a starting and an ending point (extremely unhappy/happy), yielding a continuous interval measure on which respondent marks his/her level of happiness.

- The second case study concerns the discretization of the dependent variable *mental disease* of teachers and data are a subset of the Survey on Teachers' Stress and Health (STREH) of 247 Italian teachers collected within a psychological study (Zurlo et al., 2010) about the key issues of the teachers' stress. The dependent variable is determinated by means of the sum of different tests; thus, it approaches a continuous latent variable for central limit theorem.

In both cases, the range of the continuous variable has been discretized in m classes of constant width, from $m = 3$ up to $m = 20$, and the estimated CUB models are depicted in Figure 1. Notice that the axes are scaled and do not represent the whole parameter spaces. It is evident an almost systematic increase of the level of uncertainty whereas the level of feeling may be considered stable.

To see if different scaling modifies the structure of the estimated discrete distributions $\hat{p}_j^{(m)}$, $j = 1, 2, \dots, m$, we compute location, heterogeneity and shape indexes, for each subdivision in m categories. More specifically, for each estimated CUB models, we introduce a normalized mean value (average) on $[0, 1]$ defined by $E(Y) = \frac{\sum_{j=1}^{m} j\hat{p}_j^{(m)} - 1}{m-1}$, the

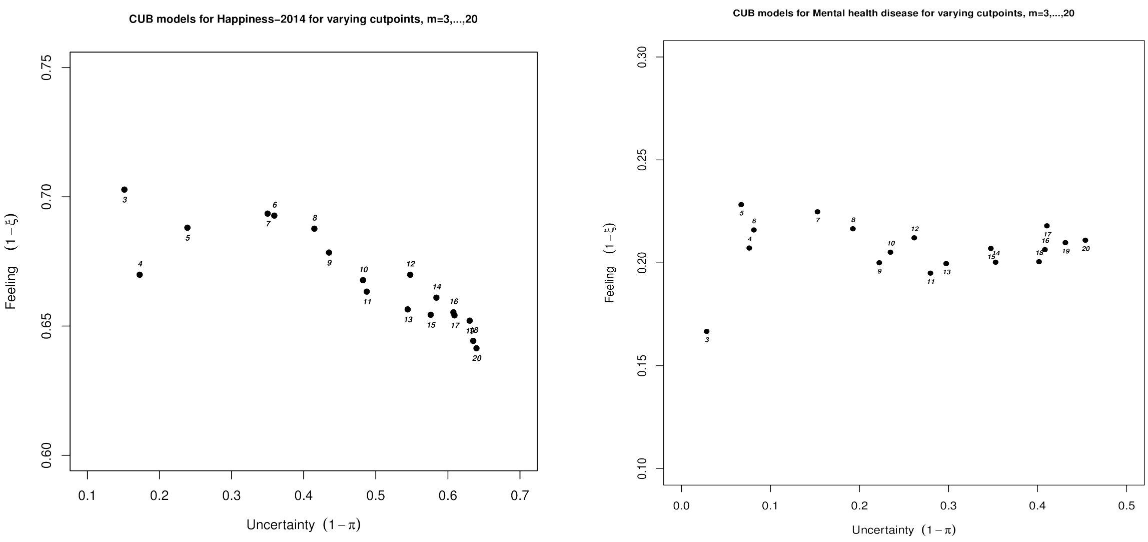

Figure 1: Effect of changing the scale on the estimated CUB models on two real data sets

already quoted heterogeneity index of Gini and the skewness and kurtosis indexes:

$$Skew(Y) = \sum_{j=1}^{m} \left[\left(\frac{j-\mu}{\sigma} \right)^3 \hat{p}_j^{(m)} \right]; \qquad Kurt(Y) = \sum_{j=1}^{m} \left[\left(\frac{j-\mu}{\sigma} \right)^4 \hat{p}_j^{(m)} \right] - 3.$$

In a sense, they capture the main futures of a discrete distribution. Figure 2 shows these measures for the selected data sets which have been analyzed by CUB models in Figure 1.

From this empirical analysis, it turns out that heterogeneity increases with m whereas the normalized average lowers.

4 A simulation experiment

A more formal approach to verify the effect of changing scale on the ordinal responses consists in simulating data from continuous distributions with different shapes, dscretize them into m classes and modifying the discrete distrbution with the inclusion of a proportion of totally random responses. Although the experiment has been conducted in several combinations of parameters for many random variables we limit ourselves to report the most significant patterns obtained in four typical cases.

In the first instance, the distributions have been selected according to the parameters specification of Table 1. In any case, $n = 1000$ observations have been generated by the latent variables over a convenient support (see last column of Table 1). Then, the sample has been discretized by cutting the range of observed values in m classes of equal size, for $m = 3, 4, \ldots, 20$. 70% of results have been mixed with 30% of a discrete

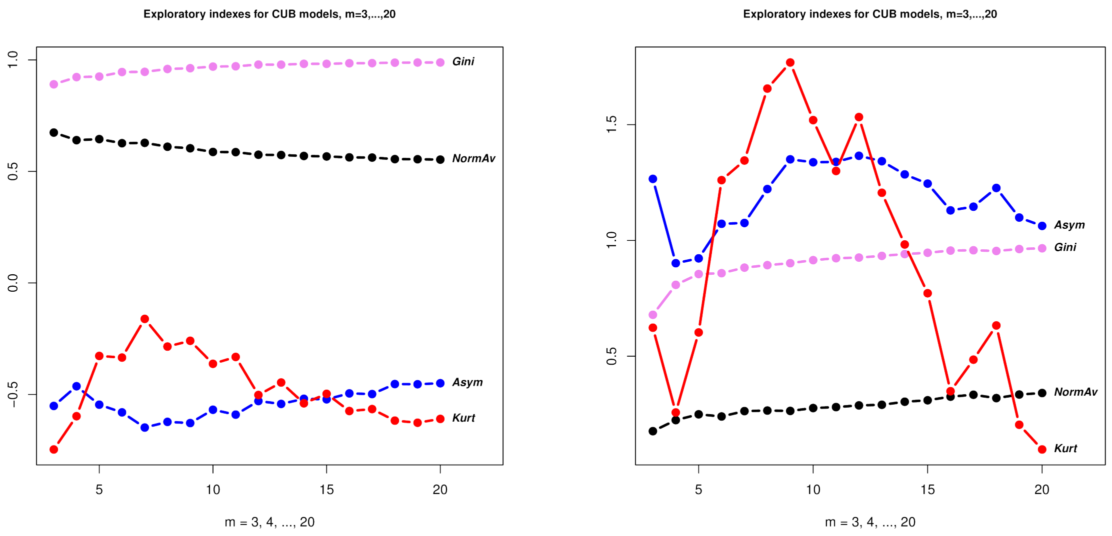

Figure 2: Main indexes of the estimated CUB models obtained by changing the scale

Uniform distribution. The obtained ordinal data have been fitted by CUB models and the estimated models are plotted in the parametric space for varying m (Figure 3).

Table 1: Selected continuous distributions for the simulation experiment

Random variables	Parameters	Skewness	Kurtosis
Gaussian	$\mu = 0, \quad \sigma^2 = 1$	0.000	0.000
Exponential	$\lambda = 0.5$	2.000	7.000
Beta	$r = 5, \quad s = 2$	-0.596	-0.120
Chi-square	$g = 5$	1.265	2.400

While it is evident the direction of the uncertainty when m increases, the skewness of the distribution and the weight of uncertainty should be considered important issues in modifying the estimated models. Thus, we first investigate right and left skewed distributions and then modify the weight of uncertainty to see this separate effect.

Figure 4 (left panel) shows the results of fitting CUB models for an increasing number of categories for two Beta random variables with $r = 5, s = 2$ and $r = 2, s = 5$, respectively. It confirms that the effect on the estimated models is almost coincident but for the feeling parameter which modifies its level (as expected). Thus, *ceteris paribus*, a modification of the skewness changes the level of the distributions. Instead, the increase of the scale does not modifies the level in any case, and thus again it turns out that the feeling parameter is nearly unaffected by a change of scale.

In Figure 4 (right panel) the effect of increasing the weight of uncertainty on a con-

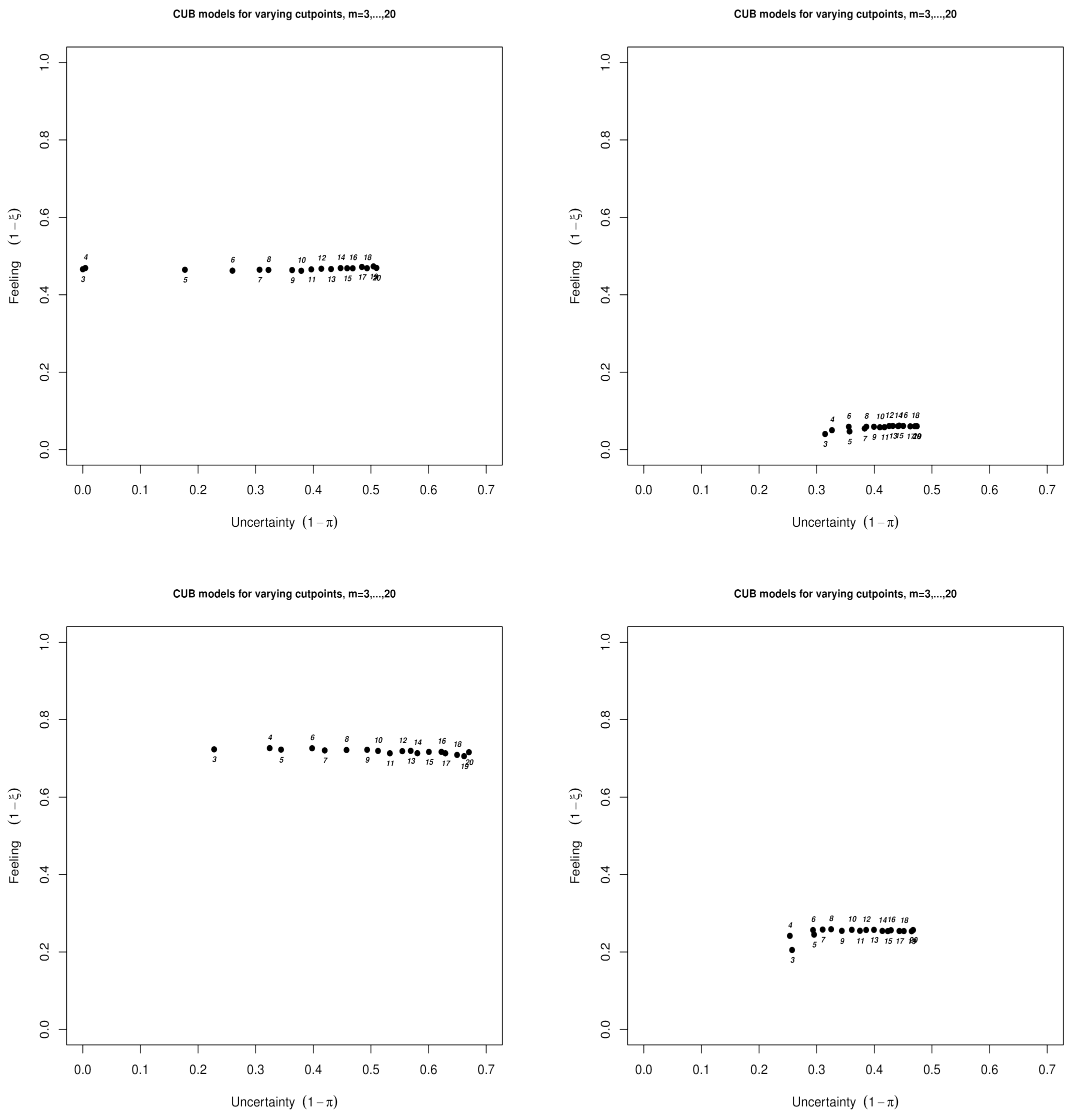

Figure 3: Effect of changing scales on different continuous distribution (Gaussian on top left, Exponential top right, Beta bottom left and Chi-square bottom right)

tinuous Beta random variable with $r = 2$, $s = 5$ is shown. The plots refer to samples where the weights of uncertainty are 0.1, 0.5, 0.9, respectively, and are represented with a line width proportional to such weights. The profiles again confirm that a modification in scale almost does not affect the feeling in any combination of uncertainty. On the contrary, uncertainty increases with the subdivision of the scale but the behaviour is more and more idiosyncratic when the weight of uncertainty is higher.

As a conclusion of the reported experiments (a subset of the largest analysis which has

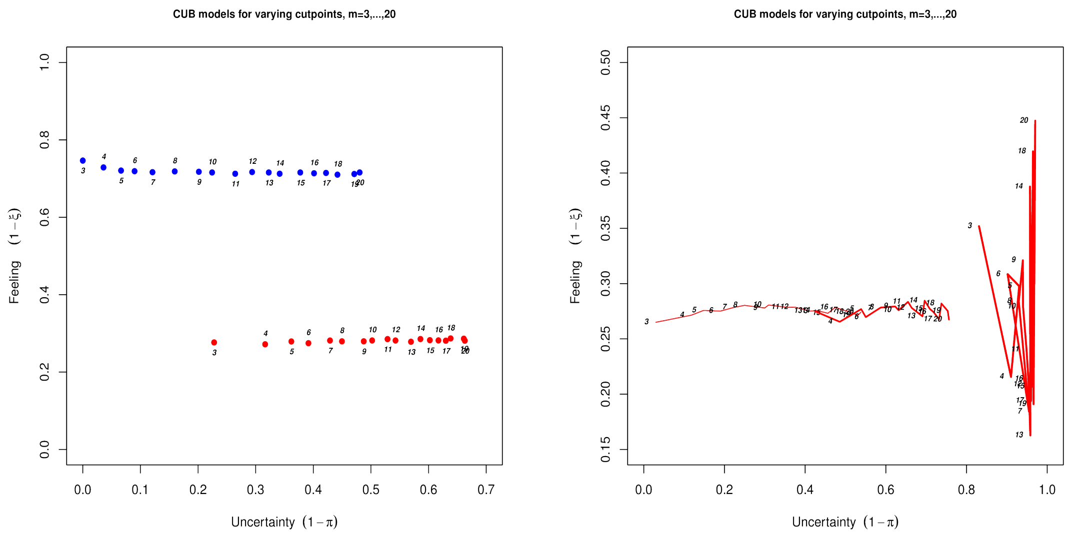

Figure 4: Effect of changing scales in case of skewness (left panel) and different weight of uncertainty (right panel)

been conducted), it turns out that the main effect of the scale modification is to increase uncertainty whereas the feeling component is substantially unaltered. The regularity of these results lowers when the weight of uncertainty in the original data set is high.

5 An approach to compare ratings with different scales

In this Section we refer to surveys where some interviewees are asked to give ordered responses by means of a m-point rating scale and other people (in different time, space, occasion) give responses to the same or similar issues on a different scale based on m^* categories.

The easiest method to compare these surveys is based on a simple proportional transformation. More specifically, people relate information on the two surveys by means of an arithmetic relationship between the ratings $(r_1, r_2, \dots, r_n)'$ and $(r_1^*, r_2^*, \dots, r_n^*)'$ based on m and m^* categories, respectively, according to:

$$r_i^* = \frac{r_i(m^*-1)+(m-m^*)}{m-1} \quad \Longleftrightarrow \quad r_i = \frac{r_i^*(m-1)+(m^*-m)}{m^*-1}, \qquad i=1,2,\dots,n.$$

Since this is a linear mapping between r_i and r_i^*, the averages of both rating scales satisfy the same relationship.

Standardization is another approach that has proved useful for evaluating empirical data (Rosenthal and Rosnow, 1991), designing experiments (Cohen, 1988), and integrating results from many studies (Hedges and Olkin, 1985).

These approaches are non-parametric and thus they does not take the features of the data generating process into account. Alternatively, if feeling and uncertainty are considered as the main components of the expressed ordinal ratings, a solution should maintain these two characteristics. More specifically, a parametric solution is suggested in the class of CUB models which allows to pass from one scale to another by considering the whole informative content of data.

As shown in previous Sections, what is really affected by a change in the scale is the uncertainty component whereas the feeling parameter may be considered as substantially stable. Then, our approach is to fit a CUB model for the given ratings $(r_1, r_2, \dots, r_n)'$ based on m points scale which produces estimates $(\hat{\pi},\, \hat{\xi})$. Assuming that the feeling parameter is generally more stationary than the uncertainty one, we pass from the probability distribution with m categories to the new one with m^* categories by selecting the π^* parameter such that the heterogeneity Gini index computed on the probability distribution of the first model be as close as possible (or coincide) to the Gini index computed with m^* categories.

Formally, if $G_m(\pi, \xi)$ denotes the Gini index (4) for a CUB random variable with m categories and specified by the parameters (π, ξ), then this approach searches for a π^* such that:

$$G_{m^*}(\pi^*, \xi) \;=\; G_m(\pi, \xi)\,.$$

Then, exploiting the relationship (5), it is immediate to solve the previous equation for π^* and obtain:

$$\pi^* \;= \pi\; \sqrt{\frac{m^*(m^*-1)}{m(m-1)}\; \frac{\displaystyle\sum_{r=1}^{m} \left[b_{r,m}(\xi)\right]^2 \;-\; 1/m}{\displaystyle\sum_{r=1}^{m^*} \left[b_{r,m^*}(\xi)\right]^2 \;-\; 1/m^*}}\;, \tag{6}$$

where $b_{r,m}(\xi)$ denotes the probability mass function of the shifted Binomial distribution over the support $\{1, 2, \dots, m\}$, and similarly for $b_{r,m^*}(\xi)$.

To get an idea of the modification in π expressed by the previous formula, we compare the values of the square root in (6) for varying ξ and for $m^* = m{+}1, m{+}2, ...$, when $m = 5$ (Figure 5). The plots show that such a ratio changes with ξ according to (approximately) a quartic function with two maxima around $\xi = 0.1$ and $\xi = 0.9$ (approximately) and with a minimum at $\xi = 0.5$. In fact, when the distribution is almost symmetric (that is around $\xi = 0.5$) the ratio is less than 1 and thus the proposed criterion reduces π, that is it increases uncertainty in the transformed distribution. Instead, for other values of ξ the effect of transformation is to increase π and thus to reduce the uncertainty which is present in the distribution. This reduction may be also very high when the values of ξ or $1 - \xi$ are located around 0.1.

For an illustrative example, let $m = 5$, $\xi = 0.1$ and $\pi = 0.4$; then, the heterogeneity Gini index is 0.9364 (Figure 6, left panel). Moving to a rating survey with $m^* = 9$, we don't change the value of the parameter ξ but, in the class of CUB models with the given $\xi = 0.1$, we will choose a CUB model such that the Gini index is as close as possible to the heterogeneity index obtained with $m = 5$. This is graphically shown in Figure 6 (middle panel) where the crossing point specifies the required π^*. In this way, the new CUB model (Figure 6, right panel) preserves the main features of the previous one with

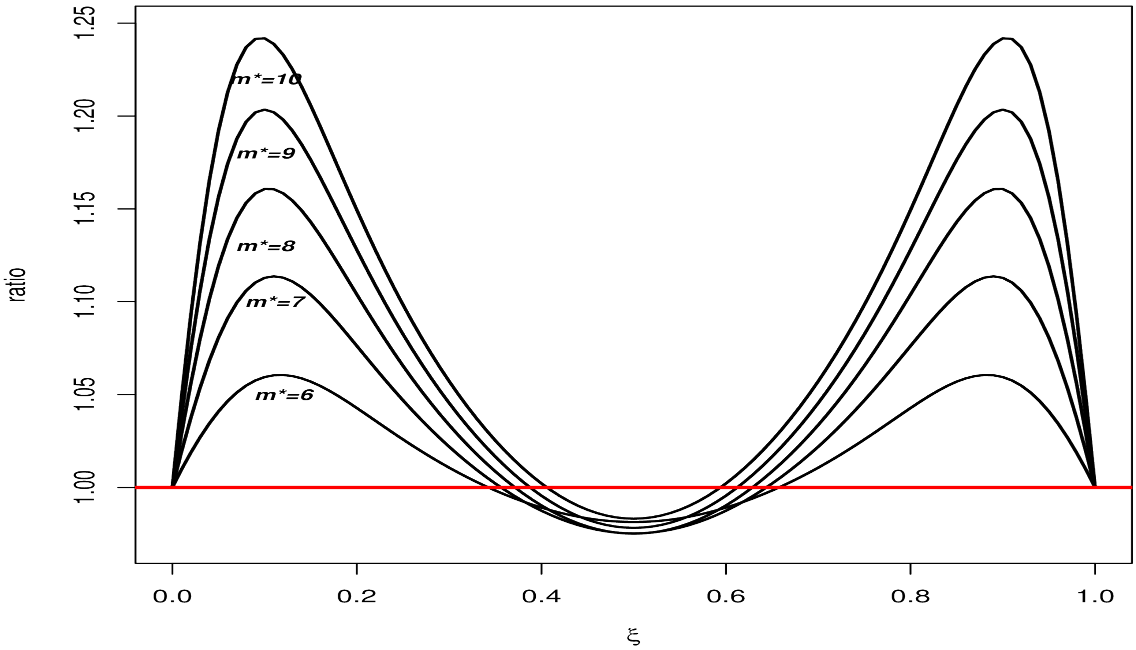

Figure 5: Ratio of formula (6) for $m = 5$ and $m^* = 6, 7, \ldots, 10$

$m^* = 9$, $\xi = 0.1$ and $\pi^* = 0.481$ and it may be used for comparative purposes. Notice that the value of π^* is increased more than 20% since it corresponds to a value of ξ which (approximately) generates the maximum ratio in formula (6), as shown in Figure 5. The global effect on the new probability mass function is a reduction of uncertainty as it is evident if one compares the distributions in the left ($m = 5$) and right ($m^* = 9$) panel of Figure 6.

Table 2: Indexes for equivalent CUB distributions

Indexes	$m = 5$	$m^* = 9$
Normalized average	0.660	0.692
Gini heterogeneity	0.936	0.936
Skewness	−0.660	−0.886
Kurtosis	−0.902	−0.496

To confirm the stability of the main features of these two distributions, Table 2 shows the main indexes computed on the CUB distributions with $m = 5$ and $m^* = 9$, respec-

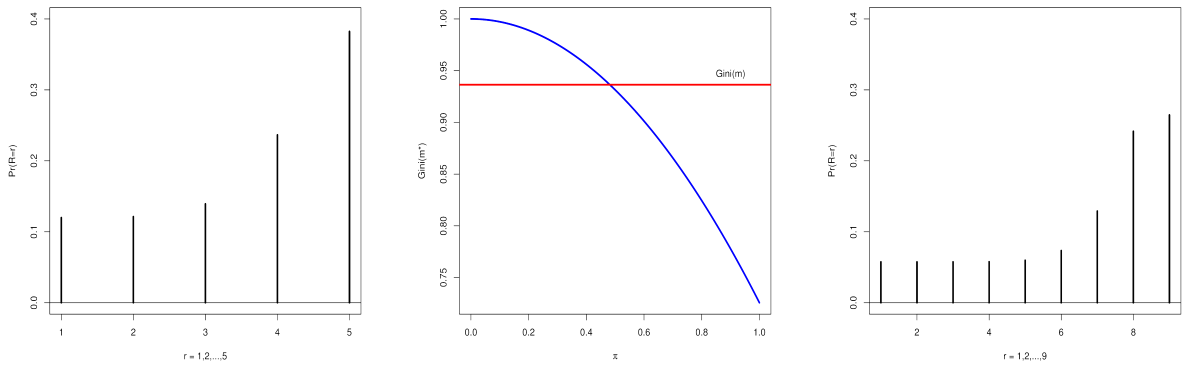

Figure 6: Equivalent CUB distributions with $m = 5$ (left panel), $m^* = 9$ (right panel) with respect to Gini index (middle panel)

tively. Thus, the proposed transformation does not modify location and heterogeneity and substantially preserves skewness and kurtosis of the modified distribution.

6 Discussion and conclusion

The framework of CUB models is useful for interpreting ordinal data in terms of parameter parsimony, adequate fitting and immediate visualization. We have exploited some of these characteristics in order to propose an approach aimed to capture the main features contained in ordinal data and to allow for the comparison of data sets originated by different surveys in similar area of investigation. The main idea is that uncertainty and feeling parameters are the fundamental information contained in the data and CUB modelling is a probability structure able to summarizes them in an effective manner.

In the previous Sections procedures have been applied both to real and simulated case studies in a context of model-based analysis. They confirm that the level of the response is substantially preserved, whereas uncertainty increases with the number of categories. An important difference between the previous studies and an empirically based approach relies on the fact that respondents react variously when faced with a different number of categories.

The empirical evidence sharply shows that the uncertainty component increases with the length of the scale. Thus, a substantive argument can be raised: "an increased number of categories induces more uncertainty in the responses" or "uncertainty is better emphasized when the number of categories increases?". A correct answer to this problem requires adequate experimental designs which we are planning for future studies.

Acknowledgements

This research has been partially supported by FIRB 2012 project (code RBFR12SHVV) at University of Perugia and by SHAPE project in the frame of STAR Programme (CUP E68C13000020003) at University of Naples Federico II, financially supported by UniNA and Compagnia di San Paolo.

References

Agresti, A. (2010). *Analysis of ordinal categorical data.* 2nd edition. Wiley, Hoboken.

Baumgartner, H. and Steenback, J. B. (2001). Response styles in marketing research: a cross-national investigation. *Journal of Marketing Research*, 38, 143–156.

Beatty, P. and Herrmann, D. (1995). A framework for evaluating don't know responses in surveys. *Proceedings of the Section for Survey Research Methods.* American Statistical Association website. Available: http://www.amstat.org/sections/SRMS/Proceedings/papers/1995-175.pdf. Accessed 2015 Jan 23.

Bradbum, N. M., Sudman, S. and Associates (1979). *Improving Interview Method and Questionnaire Design.* Jossey-Bass Publishers, San Francisco.

Buckley, J. (2009). Cross-national response styles in international educational assessments: Evidence from PISA 2006. *Technical report*, Department of Humanities and Social Sciences in the Professions Steinhardt School of Culture, Education, and Human Development. New York University.

Cicchetti, D. Y, Showalter, D. and Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of inter-rater reliability: a Monte-Carlo investigation. *Applied Psychological Measurement*, 9, 31–36.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* 2nd edition. Hillsdale, NJ: Erlbaum.

Converse, P. (1964). The nature of belief systems in mass publics. Apter, D. (eds), *Idealogy and Discontent. Free Press*, pp. 206-261.

Cox, E. P. (1980). The optimal number of response alternatives for a scale: a review. *Journal of Marketing Research.* 17, 407-422.

Cronbach, L. (1998). Response sets and test validity. *Educational and Psychological Measurement*, 6, 475–494. N. U. of Singapore, pp. 74-93, Singapore.

Cummins, R. and Gullone, E. (2000). Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. *Proceedings, Second International Conference on Quality of Life in Cities. Singapore: National University of Singapore*, N. U. of Singapore, pp. 74-93, Singapore.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50(1), 61–77.

Gambacorta, R. and Iannario, M. (2013). Measuring job satisfaction with CUB models. *Labour*, 27, 198–224.

Garner, W. R. (1960). Rating scales, discriminability and information transmission. *Psychological Review*, 67, 343–352.

Green, P. E. and Rao, Y R. (1970). Rating scales and informational recovery-how many scales and response categories to use? *Journal of Marketing*, 34, 33–39.

Gui, B. and Stanca, L. (2010). Happiness and relational goods: well-being and interpersonal relations in the economic sphere. Springer 57, pp. 105–118.

Hedges, L. Y and Olkin, L. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Iannario, M., (2009). A comparison of preliminary estimators in a class of ordinal data models. *Statistica & Applicazioni*, VII, 25–44.

Iannario, M., (2010). On the identifiability of a mixture model for ordinal data. *Metron*, LXVIII, 87–94

Iannario, M., (2012a). Modelling *shelter* choices in a class of mixture models for ordinal responses. *Statistical Methods and Applications*, 21, 1–22.

Iannario, M., (2012b). Hierarchical CUB models for ordinal variables *Communications in Statistics. Theory and Methods*, 41, 3110–3125.

Iannario, M., (2012c). Preliminary estimators for a mixture model of ordinal data. *Advances in Data Analysis and Classification*, 6, 163–184.

Iannario, M. and Piccolo, D. (2012). CUB models: statistical methods and empirical evidence. In: Kenett, R. S. and Salini, S. (eds) *Modern analysis of customer surveys: with applications using R.* Wiley, Chichester, pp. 231–258

Iannario, M. and Piccolo, D. (2014). Inference for CUB model: a program in R. *Statistica & Applicazioni. XII n.2, 177-204.*

Iannario, M. (2015). Detecting latent components in ordinal data with overdispersion by means of a mixture distribution. *Quality & Quantity, 49, 977-987.*

Iannario, M. and Piccolo, D. (2015). A generalized framework for modelling ordinal data. *Statistical Methods & Applications, (DOI) 10.1007/s10260-015-0316-9.*

Jenkins, G. D. Jr and Taber, T. D. (1977). A Monte Carlo study of factors affecting three indices of composite scale reliability. *Journal of Applied Psychology*, 62, 392–398.

King, G., Murray, C. J. L., Salomon, J. A. and Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98, 191–205.

King, G. and Wand, J. (2007). New Tools for Anchoring Vignettes. *Political Analysis*, 15, 46–66.

Kulas, J. T. and Stachowski, A. A. (2009). Middle category endorsement in oddnumbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, 43(3), 489–493.

Lietz, P. (2010). Research into questionnaire design - a summary of the literature. *International Journal of Market Research*, 52, 249–272.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55.

Lissitz, R. W. and Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60, 10–13.

Lozano, L.M., Eduardo García-Cueto, E. and Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology*, 4(2), 73–79.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society*, B, 42, 109–142.

McKelvie, S. J. (1978). Graphic rating scales: How many categories? *British Journal of Psychology*, 69, 185–202.

Manisera, M. and Zuccolotto, P. (2014). Modeling “Don't know” responses in rating scales. *Pattern Recognition Letters*, 45, 226–234.

Matell, M.S. and Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study 1: reliability and validity. *Educational and Psychological Measurement*, 31, 657–674.

Neumann, L. (1979). Effects of categorization on relationships in bivariate distributions and applications to rating scales. *Dissertation Abstracts International*, 40, 2262-B.

Neumann, J. and Neumann, Y. (1981). Comparison of six lengths of rating scales: students' attitudes toward instruction. *Psychological Reports*, 48, 399–404.

Nunnally, J. C. (1970). *Introduction to psychological measurement.* New York: McGraw-Hill.

Piccolo, D. (2003). On the moments of a mixture of uniform and shifted binomial random variables. *Quaderni di Statistica*, 5, 85–104.

Piccolo, D. (2006). Observed information matrix for MUB models. *Quaderni di Statistica*, 8, 33–78.

Piccolo, D. and D'Elia, A. (2008). A mixture model for preference data analysis. *Food Quality and Preference*, 19, 247–259.

Poe, G. S., Seeman, I., McLaughlin, J., Mehl, E. and Dietz, M. (1988). "Don't Know" boxes in factual questions in a mail questionnaire: Effects on level and quality of response. *Public Opinion Quarterly*, 52(2), 212–222.

Preston, C. C. and Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1–15.

Ramsay, J. O. (1973). The effect of number of categories in rating scales on precision of estimation of scale values. *Psychometrika*, 38, 513–533.

Rosenthal, R. and Rosnow, R. L. (1991). *Essentials of behavioral research: methods and data analysis.* 2nd edition. New York: McGraw-Hill.

Schutz, H.G. and Rucker, M.H. (1975). A comparison of variable configurations across scale lengths: an empirical study. *Educational and Psychological Measurement*, 35, 319–324.

Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. *Journal of Experimental Psychology*, 7, 456–461.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273–286.

Tutz, G. (2013). *Regression for Categorical Data*. Cambridge University Press.

Uhlaner, C. J. (1989). Relational Goods and Participation: Incorporating Sociability into a Theory of Rational Action. *Public Choice*, 62, 253–285.

Zurlo, M.C., Pes, D. and Siegrist, J. (2010). Validity and reliability of the Effort-Reward Imbalance Questionnaire in a sample of 673 Italian teachers. *International Archive of Occupational and Environmental Health*, 83, 665–674.