An empirical analysis of the coherence between fuzzy rating scale- and Likert scale-based responses to questionnaires

María Asunción Lubiano, Antonia Salas, Sara de la Rosa de Sáa,
Manuel Montenegro, and María Ángeles Gil

Abstract In dealing with questionnaires concerning satisfaction, quality perception, attitude, judgement, etc., the fuzzy rating scale has been introduced as a flexible way to respond to questionnaires' items. Designs for this type of questionnaires are often based on Likert scales. This paper aims to examine three different real-life examples in which respondents have been allowed to doubly answer: in accordance with either a fuzzy rating scale or a Likert one. By considering a minimum distance-based criterion, each of the fuzzy rating scale answers is associated with one of the Likert scale labels. The percentages of coincidences between the two responses in the double answer are computed by the criterion-based association. Some empirical conclusions are drawn from the computation of such percentages.

1 Introduction

In designing questionnaires concerning variables which cannot be measured by means of exact numerical values but can be graded to some extent (as it happens with satisfaction, quality perception, agreement level, and so on), commonly employed scales are Likert ones. Items in Likert scale-based questionnaires are responded by choosing among a list of a few pre-specified answers the one that best represents respondent's valuation, rating, opinion, etc. Likert scale-based answers can be usually ordered with respect to a certain criterion (say degree of satisfaction, degree of goodness, degree of agreement, etc.).

Hesketh *et al.* [5] (see also Hesketh and Hesketh [4]) proposed the so-called fuzzy rating scale to allow a complete freedom and expressiveness in

Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, 33071 Oviedo
`lubiano@uniovi.es,antonia@uniovi.es,rosasara@uniovi.es,mmontenegro@uniovi.es,`
`magil@uniovi.es`

responding, without respondents being constrained to choose among a few pre-specified responses. By drawing the fuzzy number that best represents respondent's valuation, the fuzzy rating scale captures the logical imprecision associated with such variables and allows us to have a rich continuous scale of measurement. In this way, the fuzzy rating scale somehow combines the power of the fuzzy linguistic scales with the power of visual analogue scales.

In previous papers, responses to items in synthetic and real-life questionnaires based both on Likert and fuzzy rating scales have been empirically compared by means of different statistical tools (see, for instance, De la Rosa de Sáa *et al.* [1], Gil *et al.* [3] and Lubiano *et al.* [7]).

Since responses in accordance with the two scales are collected in a linked way (i.e., respondents supply a double answer), one question that arises is whether or not respondents follow a kind of systematic classification of the fuzzy rating scale-based responses into classes that could be identified with Likert's possible answers.

This paper aims to examine this question by analyzing three real-life examples involving questionnaires with double response type items. For this purpose a criterion based on a distance between Likert and fuzzy responses (actually, between numerically encoded Likert and fuzzy responses) is applied. This analysis evidences that the coincidences between the expected Likert response and the one really chosen are high, but up to 90%. This suggests that in assigning fuzzy rating scale-based responses people behave in a very free way, without trying to exactly follow a kind of fuzzy linguistic description of a Likert response. Furthermore, this fact corroborates to some extent that, as it has been frequently pointed out in the literature, the usual numerical encoding of Likert responses is not appropriate enough.

2 Preliminaries

Fuzzy numbers are often considered to express imprecise data because of their ability and power to precisiate the imprecision and to be mathematically handled.

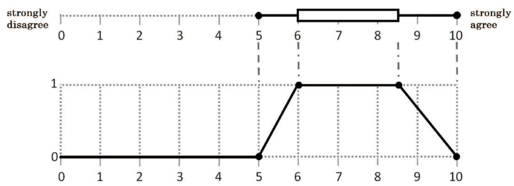
Definition 1. A mapping $\widetilde{U} : \mathbb{R} \to [0,1]$ is said to be a (bounded) *fuzzy number* if its α-levels

$$\widetilde{U}_\alpha = \begin{cases} \{x \in \mathbb{R} : \widetilde{U}(x) \geq \alpha\} & \text{if } \alpha \in (0,1] \\ \mathrm{cl}\{x \in \mathbb{R} : \widetilde{U}(x) > 0\} & \text{if } \alpha = 0 \end{cases}$$

(with cl denoting the topological closure) are nonempty compact intervals for all $\alpha \in [0,1]$. The class of (bounded) fuzzy numbers will be denoted by $\mathcal{F}_c^*(\mathbb{R})$.

In accordance with Hesketh *et al.* [5] (see also Hesketh and Hesketh [4]), the guideline for the use of fuzzy numbers through the so-called *fuzzy rating scale* is the following:

1. A reference bounded interval/segment $[a, b]$ is first considered. This is often chosen to be $[0, 10]$ or $[0, 100]$, but the choice of the interval is not at all a constraint. The end-points are often labeled in accordance with their meaning referring to the degree of satisfaction, quality, agreement, and so on.
2. The *core*, or 1-level set, associated with the response is determined. It corresponds to the interval consisting of the real values within the reference one which are considered to be as 'fully compatible' with the response.
3. The *support*, or its closure or 0-level set, associated with the response is determined. It corresponds to the interval consisting of the real values within the referential that are considered to be as 'compatible to some extent' with the response.
4. The two preceding intervals are 'linearly interpolated' to get a trapezoidal fuzzy number.



In accordance with Likert scales, people respond to items by specifying their feeling with respect to a statement on a symmetric 'agree-disagree', or 'extremely high-extremely low', etc., scale. This specification is performed by choosing one among several given points representing some key degrees of agreement/suitability, etc. To analyze Likert scale-based responses, such points are encoded by means of consecutive integer numbers.

The question posed in Section 1, about whether or not fuzzy rating scale-based responses could be into k-point Likert's ones, is to be answered in this paper by considering the distance-based mapping $\iota : \mathcal{F}_c^*(\mathbb{R}) \to [a, b]_k = \{a, a+(b-a)/(k-1), \ldots, a+(b-a)(k-2)/(k-1), b\}$ (with $[a, b] =$ reference interval, so that the integer consecutive codes have been re-scaled in accordance with the reference interval) such that $\widetilde{U} \mapsto \arg\min_{i\in[a,b]_k} \rho_2(\widetilde{U}, \mathbb{1}_{\{i\}})$, that is,

$$\iota(\widetilde{U}) = \arg \min_{i\in[a,b]_k} \sqrt{\int_{[0,1]} \frac{(\inf \widetilde{U}_\alpha - i)^2 + (\sup \widetilde{U}_\alpha - i)^2}{2} \, d\alpha},$$

ρ_2 being the well-known L^2 metric introduced by Diamond and Kloeden [2].

3 Real-life examples

In this section, we are going to examine three real-life situations in which questionnaires allowing to choose-draw double Likert type-fuzzy rating type responses have been conducted. In each of the examples, we have determined the percentages of coincidences between the expected Likert response (more concretely, the image of the fuzzy rating response through ι and the assessed Likert response).

Example 1. By using an online computerized application an experiment has been performed in which people have been asked for their perception of the relative length of different line segments with respect to a pattern longer one (see http://bellman.ciencias.uniovi.es/SMIRE/Perceptions.html).

On the center top of the screen the longest (pattern) line segment has been drawn in black. This segment is fixed for all the trials, so that there is always a reference for the maximum length. At each trial a grey shorter line segment is generated and placed below the pattern one, parallel and without considering a concrete location (i.e., indenting or centering). For each respondent, line segments are generated at random, although to avoid the variation in the perception of different respondents can be mainly due to the variation in length of different generated segments, the (27 first) trials for two respondents refer to the same segments but appearing in different position and location.

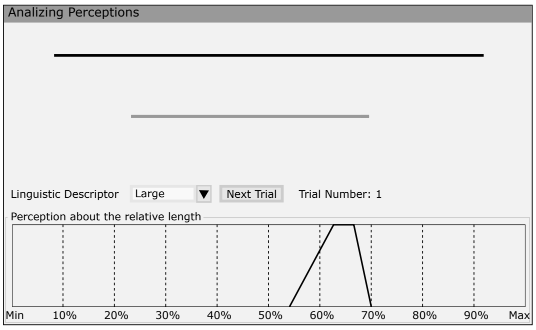


Fig. 1 Example of a double response from the computerized application in Example 1

The computerized application explains the formalization and meaning of the fuzzy rating values (see Figure 1), with reference interval $[0, 100]$. People have participated online by providing with their judgement of relative length for each of several line segments. Each of these judgements can be doubly expressed: by choosing a label from a 5-point Likert-like list (0 = VERY SMALL, 25 = SMALL, 50 = MEDIUM, 75 = LARGE, 100 = VERY LARGE), and by using the fuzzy rating method.

25 respondents (all with a university scientific background) have supplied 1387 double responses after the corresponding trials. The dataset can

be found in http://bellman.ciencias.uniovi.es/smire/Archivos/Perceptions dataset.pdf. The percentage of coincidences through the minimum distance criterion equals 84.93%.

Example 2. A sample of 70 people with different age, background and work type and position has been considered to fill a restaurant customer satisfaction questionnaire with 14 items by using a double response-type form (see http://bellman.ciencias.uniovi.es/smire/FuzzyRatingScaleQuestionnaire-Restaurants.html).

The questionnaire has been conducted by a few students of a Master on Soft Computing and Intelligent Data Analysis held in Mieres in 2011-2012. Figure 2 displays the excerpt of the form to be filled corresponding to one of the involved items.

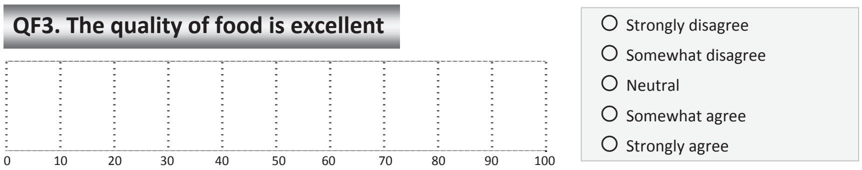


Fig. 2 Excerpt of a questionnaire about the satisfaction with the quality of restaurants in Example 2

The form allows the double response, where Likert-like ones are chosen from a 5-point Likert scale ($0 =$ STRONGLY DISAGREE, $25 =$ SOMEWHAT DISAGREE, $50 =$ NEUTRAL, $75 =$ SOMEWHAT AGREE, $100 =$ STRONGLY AGREE) and the fuzzy ones have reference interval $[0, 100]$.

The dataset with 980 double responses can be also found in the webpage including the form. The percentage of coincidences through the minimum distance criterion equals 78.16%.

Example 3. This third example is related to the well-known questionnaire TIMSS-PIRLS 2011 which is conducted on populations of (nine to ten years old) fourth grade students and concerns their opinion and feeling on aspects regarding reading, math, and science. This questionnaire is rather standard and most of the involved questions have to be answered according to a 4-point Likert scale ($0 =$ DISAGREE A LOT, $10/3 =$ DISAGREE A LITTLE, $20/3 =$ AGREE A LITTLE, $10 =$ AGREE A LOT).

The original questionnaire form has been adapted to allow a double-type response, the original Likert and a fuzzy rating scale-based one with reference interval $[0, 10]$ (see Figure 3 for one of the items, and the webpage http://bellman.ciencias.uniovi.es/SMIRE/FuzzyRatingScaleQuestionnaire-SanIgnacio.html for the full paper-and-pencil and computerized forms and datasets).

As a differential feature and to ease the relationship between the two scales for respondents, each numerically encoded Likert response has been superimposed upon the reference interval of the fuzzy rating scale part.

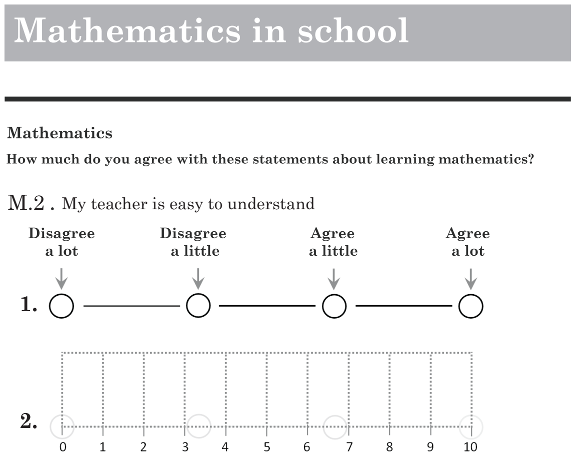


Fig. 3 Example of the double-response form to a question in Example 3

The questionnaire involving these double-response questions has been conducted on 69 fourth grade students from Colegio San Ignacio (Oviedo-Asturias, Spain). The dataset with 599 double responses can be also found in the webpage including the form. The percentage of coincidences through the minimum distance criterion equals 81.47%.

The above indicated percentages for the three examples have been also computed with some other few metrics, even some ones assessing different weights to different α-levels (more concretely, assessing weights so that the higher the α the higher/lower the weight). Percentages have been scarcely affected by the choice of the metric.

4 Some remarks from the analysis of the real-life examples

As a summary of the analysis of the percentages in the three examples in Section 3 we can empirically conclude that background, age and sample sizes seem not to be very influential, as we could formerly suspect. Actually, we should confess that children in the third example, which are much younger and are assumed not to have yet a high background, have positively surprised us with their ease to catch the idea in just 15 minutes of explanation.

On the other hand, we can also conclude that in real-life people having the opportunity of the double response is not necessarily guided by what Likert labels can mean. In fact, it seems that people take advantage of the flexibility, freedom and expressiveness of the fuzzy rating scale to draw their valuations and they make it rather independently of their Likert assessment even in case they have to do it simultaneously. This corroborates what has been statistically concluded by Lubiano *et al.* [6, 7]: Likert scales 'aggregate' in some sense valuations which could be 'precisiated' through fuzzy numbers, so relevant information can be lost when using Likert scales.

This paper also adds that the real-life aggregation does not correspond in practice to a natural (distance-based) partition of the fuzzy rating scale-based responses. And, probably, there is no criterion which could properly mimic human association. In this way, the following responses have been taken from the dataset of the responses in Example 3 to the item $M.2$ in Figure 3, namely, "My math teacher is easy to understand". Figure 4 shows two very different fuzzy responses to this item for which both the distance-based association and the real choice from the 4-point Likert scale coincide (disagree a little). Figure 5 shows two rather close fuzzy responses to this item for which the distance-based association from the 4-point Likert scale coincide (agree a little), but the real choices do not.

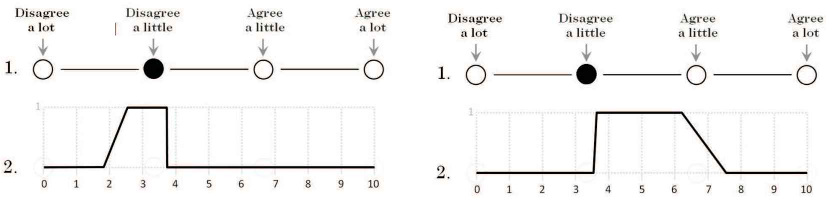


Fig. 4 Example of two fuzzy responses from Example 3 for which both the real and the minimum distance-based Likert labels coincide

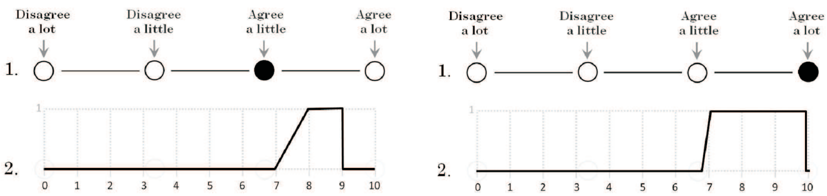


Fig. 5 Example of two fuzzy responses from Example 3 for which the minimum distance-based Likert labels coincide, but the real choices do not

To end this paper, we would like simply illustrating these conclusions with a simple instance also taken from the dataset of the responses in Example 3 to the item $M.2$ in Figure 3. Among the 69 double responses to this item, 10 of the Likert components have not matched with the minimum distance Likert (that we can refer to as the expected Likert label) These responses have been gathered in Table 1, where we can easily see that 8 of them correspond to the 8 widest (w.r.t. support, and, mostly, w.r.t. core) fuzzy responses, whereas the other 2 correspond to narrower fuzzy responses but showing close distances (w.r.t. the maximum distance 10) to two of the encoded Likert responses.

Finally, it should be emphasized that the high percentage of coincidences of the real and the minimum distance-based 'Likertization' processes should not be viewed as an argument in favour of the use of the Likert scale in

Table 1 Responses to the item "My math teacher is easy to understand" in Example 3 for which the real 4-point Likert choice and the minimum distance one do not match

inf $\widetilde{U}_0$	inf $\widetilde{U}_1$	sup $\widetilde{U}_1$	sup $\widetilde{U}_0$	Chosen Likert	dist to 0	dist to 10/3	dist to 20/3	dist to 10	Mindist Likert ι	width support	width core
3.5	3.55	6.25	7.5	10/3	5.47	2.52	2.24	5.09	20/3	**4**	2.7
5.95	6	9.2	10	10	8	4.81	2.14	2.86	20/3	**4.05**	3.2
4.9	4.9	8.45	9.975	10	7.38	4.31	2.21	3.66	20/3	**5.075**	3.55
8	8.5	8.5	9	20/3	8.5	5.17	**1.86**	**1.53**	10	1	0
3.4	4.825	9.95	9.95	10	7.62	4.72	2.96	4.17	20/3	**6.55**	5.125
3.175	5,025	7.5	9,95	10	6.85	3.9	2.41	4.31	20/3	**6.775**	2.475
8	8.5	9.2	9.2	20/3	8.74	5.41	**2.11**	**1.36**	10	1.2	0.7
5.6	6.7	9.15	10	10	8.05	4.85	2.11	2.75	20/3	**4.4**	2.45
5.825	5.85	9.875	9,95	10	8.13	4.98	2.37	2.94	20/3	**4.125**	4.025
2.5	4.625	4.625	6.9	20/3	4.83	1.84	2.37	5.49	10/3	**4.4**	0

contrast to the fuzzy rating one. On the contrary, situations like those in Figures 4 and 5 clearly illustrate the need for the last scale, whenever it can be properly employed and data are to be statistically analyzed. Thus, the statistical analysis of the Likert responses in Figure 4 doesn't distinguish between them, whereas the responses are indisputably different if the fuzzy rating scale is considered. Consequently, many errors, deviations, differences, are often neglected in using Likert scales.

Acknowledgements Authors are grateful to Colegio San Ignacio in Oviedo-Asturias (Spain) for allowing us to collect the data in the real-life example. The research in this paper has been partially supported by/benefited from Principality of Asturias Grants GRUPIN14-101 and Severo Ochoa BP12012 (De la Rosa de Sáa), and the Spanish Ministry of Economy and Competitiveness Grants MTM2015-63971-P and MTM2013-44212-P. Their financial support is gratefully acknowledged.

References

1. De la Rosa de Sáa S, Gil MA, González-Rodríguez G, López MT, Lubiano MA (2015) Fuzzy rating scale-based questionnaires and their statistical analysis. IEEE T Fuzzy Syst 23:111–126
2. Diamond P, Kloeden P (1999) Metric spaces of fuzzy sets. Fuzzy Sets Syst 100:63–71
3. Gil MA, Lubiano MA, De la Rosa de Sáa S, Sinova B (2015) Analyzing data from a fuzzy rating scale-based questionnaire. A case study. Psicothema 27:182–191
4. Hesketh T, Hesketh B (1994) Computerized fuzzy ratings: the concept of a fuzzy class. Behav Res Meth Ins C 26:272–277
5. Hesketh T, Pryor R, Hesketh B (1988) An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences. Int J Man-Mach Stud 29:21–35
6. Lubiano MA, De la Rosa de Sáa S, Montenegro M, Sinova B, Gil MA (2016) Descriptive analysis of responses to items in questionnaires. Why not using a fuzzy rating scale? Inform Sci (submitted, under review)
7. Lubiano MA, Montenegro M, Sinova B, De la Rosa de Sáa S, Gil MA (2016) Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications. Eur J Oper Res 251:918–929