

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/344688166>

Fuzzy rating scales: Does internal consistency of a measurement scale benefit from coping with imprecision and individual differences in psychological rating?

Article in Information Sciences · October 2020

DOI: 10.1016/j.ins.2020.10.042

CITATIONS

0

READS

264

3 authors:



María Asunción Lubiano

University of Oviedo

61 PUBLICATIONS 922 CITATIONS

SEE PROFILE



Antonio L. Garcia-Izquierdo

University of Oviedo

61 PUBLICATIONS 448 CITATIONS

SEE PROFILE



María Angeles Gil

University of Oviedo

231 PUBLICATIONS 3,528 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Cátedra Asturias Prevención [View project](#)



Evaluation and Selection of Public Managers: equality measures and procedures [View project](#)

Fuzzy rating scales: does internal consistency
of a measurement scale benefit
from coping with imprecision and individual differences
in psychological rating?^{*,**}

María Asunción Lubiano^a, Antonio García-Izquierdo^b, María Ángeles Gil^a

^a*Departamento de Estadística, I.O. y D.M., Facultad de Ciencias
Universidad de Oviedo, E-33071 Oviedo, Spain*

^b*Departamento de Psicología, Facultad de Psicología
Universidad de Oviedo, E-33071 Oviedo, Spain*

Abstract

Measuring psychological variables (attitudes, opinions, perceptions, feelings, etc.) there is a need for rating scales coping with both the natural imprecision and individual differences. In this respect, the so-called fuzzy rating scales have been introduced as a doubly continuous instrument allowing to capture both imprecision and individual differences. Aiming to show the advantages of using fuzzy rating scales in the setting of questionnaires, the extended Cronbach α is considered to quantify the internal consistency associated with constructs involving fuzzy rating scale-based items. This extended tool allows us to draw interesting conclusions, the main one supporting the use of fuzzy rating scales instead of standard ones (namely, Likert type, visual analogue, and even fuzzy linguistic scales). Although general theoretical conclusions could not be drawn, unequivocal majority trends can be stated from simulation-based and real-life examples.

Keywords: extended α , fuzzy linguistic scales, fuzzy rating scales, Likert scales, psychological variables, simulated fuzzy data, visual analogue scales

1. Introduction

In introducing Fuzzy Sets and Logic [92, 93, 94] Zadeh anticipated that they would play an important role in human thinking. In particular, appli-

^{*}The paper is dedicated to the memory of our admired friend Dr. Enrique H. Ruspini.

^{**}This research has been partially supported by the Principality of Asturias/FEDER Grant GRUPIN-IDI2018-000132 and the Spanish Ministry of Science, Innovation and Universities Grants PSI2013-44854-R and MTM2015-63971-P and the Spanish Ministry of Science and Innovation PID2019-104486GB-I00. Their support is gratefully acknowledged.

Email addresses: lubiano@uniovi.es (María Asunción Lubiano), angarcia@uniovi.es (Antonio García-Izquierdo), magil@uniovi.es (corresponding author) (María Ángeles Gil)

cations of fuzzy approaches and fuzzy-based techniques to Psychology were foreseen and developed since the very beginning (see, for instance, [40], or the edited book [97]). And they are more active and challenging nowadays than in the past (see, for instance, [71, 72, 75, 83] for recent reviews about).

Psychological measurement has traditionally dealt with giving or choosing either numbers, linguistic/qualitative labels or symbols to a wide array of variables. Psychological dimensions can be measured by different approaches, but they have relied mainly on questionnaires. As measurement is intended to be difficult into capturing all of the intrinsic nuances of psychological dimensions, most questionnaires have been preferred to including several items for every dimension, attempting to achieve a higher chance of respondents to be consistent.

As it has been recently highlighted [87], the choice of response format should play a crucial role in designing a new questionnaire. In this respect, choosing the response format should be an explicit step to be added to other well-known steps such as the construct definition, the target population, the item selection, and so on. The most common response format in questionnaires is the rating scale response format. A recent guest editorial of the special issue on ‘Design aspects of rating scales in questionnaires’ [56] points out that “... Since their introduction by Thurstone [79] and Likert [44], rating scales have been determinant in questionnaires. A rating scale usually defines the graduations out of a continuum such as agreement, intensity, frequency, or satisfaction. Respondents evaluate questions and items by marking the appropriate category, which usually concerns personal characteristics, opinions, and behavior.”

When aspects to be studied through questionnaires can be numerically measured in a direct way, the corresponding items make use of numerical scales and the statistical analysis of the associated responses can be analyzed by means of different well-known statistical techniques. Nevertheless, many aspects of interest are not directly measurable, so other rating scales should be considered.

Psychological measurement can be frequently seen as how much individuals agree or disagree with each of a series of statements related with psychological latent variables. In measuring these variables several rating scales have been considered in the literature. Among these scales, Likert type and visual analogue (this last one also coined as graphic or continuous rating) are probably the most popular ones (e.g., [24, 37, 43, 66, 77, 80]).

Likert Scales format [44] consists of some scores indicating the strength of the agreement with several assertions, the Likert type items. Sometimes, these numbers are combined with or replaced by linguistic expres-

sions which are, for instance, adverbs of frequency and quantity. Despite Likert scales have been adopted for many social science research communities, they present some controversy and debate concerning several issues, among them, the nature of the response categories and the uses of the scores. Likert scale-based items in a questionnaire are usually easy to respond, and the answers (before numerically ‘encoding’ them) can partially capture a certain involved imprecision.

Since the choice is made within a list of a few possible ones, anchored for the Likert options, individual differences are almost systematically overlooked. Furthermore, the number of applicable (non-parametric) techniques to statistically analyze Likert data is quite limited and relevant statistical information is usually lost in the analysis (see [86] for a recent short discussion about).

Visual Analogue Scales (VAS’s for short) were mostly considered to overcome the limitations with ordinal discrete Likert-type scales. VAS’s have a long tradition in psychological measurement, and were introduced [28] in attaining judgements about employees by supervisors (see [90] for a recent historical study). Respondents to a VAS item/scale, mark their level of agreement to a statement by indicating a position along a continuous line between two end-points, permitting an infinite number of gradations.

In contrast to Likert scales, a VAS properly captures individual differences because the choice is made within a continuum of possible options (actually, a bounded interval). However, the choice of the single point that best represents rater’s score in visual analogue scales is frequently neither easy nor natural in the usual imprecise settings (see [85]). To require a full precision seems rather unrealistic in connection with most of psychological variables.

Aiming to appropriately capture the natural imprecision involved in measuring most of psychological variables, fuzzy scales have been introduced. The best known fuzzy scales of measurement are those associated with fuzzy linguistic variables, which can mathematically cope with the imprecision of linguistic labels (see [94]). But to also cope with the individual differences, which are statistically relevant, the fuzzy rating scales (see [33, 34]) are a very appropriate instrument.

In Section 2 the preliminaries for fuzzy scales as well as the two main types of fuzzy scales are recalled. Furthermore, a few comments from some comparative studies (e.g., [15, 26, 47, 50, 51]) are given to state that fuzzy scales could lead to statistical conclusions different from those with Likert’s and VAS’s.

Since fuzzy rating scales are the only ones capturing simultaneously imprecision and individual differences in dealing with questionnaires, they look to be more informative from a statistical perspective than the Likert type, the visual analogue and the fuzzy linguistic scales. This paper presents for the first time a study allowing us to prove such an intuition by considering a unique ‘ranking’ tool. This tool is the internal consistency of the involved constructs, which is to be based on the extension of the well-known Cronbach α coefficient and is presented in Section 3.

Three key research questions will be posed about the extended tool in order to state that

- the shape of fuzzy responses scarcely affects the value of the extended Cronbach coefficient (STUDY 1, Section 4);
- the usual behavior of Cronbach’s coefficient in connection with the number of items of the constructs for numerical or numerically encoded responses is preserved when fuzzy scales are considered (STUDY 2, Section 5);
- fuzzy rating scales are the ones very mostly showing the highest internal consistency (STUDY 3, Section 6).

It should be highlighted that there are no general conclusions concerning the posed questions, so no theoretical results can be established for this purpose. However, by designing adequate novel simulation studies mimicking real-life situations, majority trends can be formally obtained. Therefore, STUDIES 1–3 will be carried out on the basis of such simulations. Finally, a real-life example is examined to complete these simulation-based studies and corroborate the majority conclusions.

2. Fuzzy scales for psychological measurement

Fuzzy scales make use of a particular type of fuzzy set that model imprecise amounts: the fuzzy numbers.

Definition 2.1. A (bounded) **fuzzy number** is an imprecise amount or quantity that is formally characterized by means of a mapping $\tilde{U} : \mathbb{R} \rightarrow [0, 1]$ (\mathbb{R} denoting the space of real numbers), where $\tilde{U}(x)$ is interpreted as the degree to which real number x is compatible with \tilde{U} , and such that for each $v \in (0, 1]$ the v -**level set**, $\tilde{U}_v =$ set of real numbers which are compatible with to a extent at least equal to v , is a nonempty closed and bounded interval, $[\inf \tilde{U}_v, \sup \tilde{U}_v]$, and the **support set**, $\text{supp}(\tilde{U}) =$ set of real numbers which are compatible to some extent with \tilde{U} , is a bounded interval. The 0-level $\tilde{U}_0 = [\inf \tilde{U}_0, \sup \tilde{U}_0]$ is defined to be the mathematical closure of $\text{supp}(\tilde{U})$.

A well-known and easy-to-use and easy-to-interpret type of fuzzy numbers are the trapezoidal ones.

Definition 2.2. If $a, b, c, d \in \mathbb{R}$ and $a \leq b \leq c \leq d$, the **trapezoidal fuzzy number** $\text{Tra}(a, b, c, d)$ is given for all $v \in [0, 1]$ by

$$(\text{Tra}(a, b, c, d))_v = [a + v(b - a), d + v(c - d)],$$

and is graphically displayed in Figure 1.

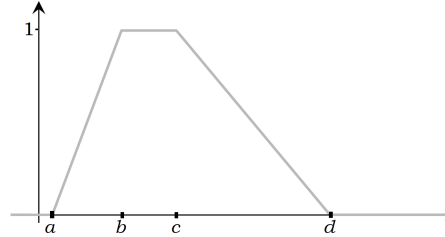


Figure 1: Graphical representation of the trapezoidal fuzzy number $\text{Tra}(a, b, c, d)$

Real numbers can also be viewed as a special type of fuzzy number, by identifying them with the indicator function of the corresponding singleton.

The concept of *fuzzy scale* was introduced to add properties to nominal and ordinal scales, by establishing a bridge between strongly defined measurements, as the visual analogue, and weakly defined measurements as the Likert-type (see [4]). Fuzzy scales are progressively being applied to social sciences to address different issues, like behavior-based safety management [14], typologies [21], human behavior analysis [1], employee trust [22], job satisfaction [19], sentiment analysis [39, 98], personal characteristics by using fuzzy set qualitative comparative analysis-fsQCA [55, 69], evaluation [25], and so on.

The main fuzzy scales are the so-called fuzzy linguistic and fuzzy rating, which we are going to explain now.

2.1. Fuzzy linguistic scales

Fuzzy Linguistic Scales (FLS's), associated with the so-called fuzzy linguistic variables, were stated by Zadeh [94] as a flexible alternative to the numerical encoding of Likert and semantic differential scales [62]. In fact, the numerical encoding does not take into account the essential imprecision accompanying 'values' of most of psychological variables, whereas the fuzzy linguistic encoding does. So, in these studies, the fuzzy-valued encoding of Likert options seems to be more appropriate fuzzy quantifiers (see [95]) and it has been especially applied in the framework of personnel selection (see, for instance, [18, 41]).

The imprecision associated with most of linguistic responses can be suitably grasped by means of fuzzy linguistic scales, but individual differences cannot. More concretely, as already has been pointed out before, respondents/raters choosing the same underlying Likert response to a given question can have somewhat different views or opinions; and respondents/raters having close positions can choose different Likert responses.

2.2. Fuzzy rating scales

To overcome the last drawback FRS's emerged in the course of the study of problems related to psychological research. Hesketh *et al.* [33, 34] introduced **Fuzzy Rating Scales** (FRS) as an extension of both the semantic differential and the visual analogue scales. FRS's (also called computerized fuzzy graphic rating scales) offer a way of addressing both the range/position of opinions along with (possibly assymmetric) latitudes of acceptance. Traditional mathematics failed in capturing such a complexity, whereas fuzzy sets could handle it as it is now summarized.

Definition 2.3. *A (computerized) **fuzzy rating scale** (or computerized fuzzy graphic rating scale) allows a rater to draw the fuzzy number that best represents rater's score as follows (see Figure 2):*

- Step 1.** *a reference bounded interval (often with anchored extremes), $[l_0, u_0]$, is first considered; there is no constraint about the choice, although intervals like $[0, 10]$ or $[0, 100]$ are very usual;*
- Step 2.** *the core/1-level corresponds to the interval of real numbers which are fully compatible with rater's score;*
- Step 3.** *the 0-level is the mathematical closure of the associated support, which corresponds to the interval of real numbers which are compatible to some extent with rater's score;*
- Step 4.** *these two intervals are 'interpolated' to get fuzzy rater's score; if this 'interpolation' is considered to be linear, then the fuzzy score will be given by a trapezoidal fuzzy number.*

Remark 2.1. At this point, it should be clarified that to consider a linear 'interpolation' in **Step 4** is not a must (see [34]). Actually, most of the already developed statistical methods could be carried out for any type of fuzzy numbers, although computations would become usually much more cumbersome. However, and because of the intuitive interpretation of the core and support of a fuzzy score in an FRS, the linear interpolation would certainly ease the training of respondents to a questionnaire involving FRS-based items. Furthermore, the trapezoidal shape allows simple computerized

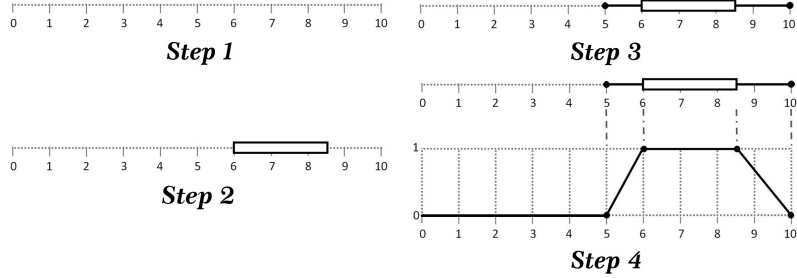


Figure 2: Steps in establishing a fuzzy score on the basis of a certain FRS

designs for FRS-based items, and computerized forms avoid that meaningless responses can be given. Hesketh *et al.* [33] highlighted that the trapezoidal version of the FRS “... achieves simplicity of presentation to a respondent at the price of an assumption about the shape of the fuzzy function generated...” Along this paper it will be proved that, as shown for other summary measures, the extended α is also scarcely affected by the chosen shape for the fuzzy rating score. Consequently, the referred ‘price of the assumption about the shape’ is rather neglectable.

Because of their flexibility, FRS’s can cope to a full extent with both: the intrinsic imprecision associated with human thought involved in the rating of psychological variables, and the individual differences associated with such a rating. More concretely, FRS’s are doubly (horizontally and vertically) continuous and they *capture both the inherent imprecision* (through the horizontal view for fixed values on the ordinate axis) *and the individual differences* (through the vertical view for fixed values on the abscissa axis). As quoted from Zadeh [96] “paradoxically, one of the principal contributions of fuzzy logic - a contribution which is widely unrecognized - is its high power of precisiation of what is imprecise.” Consequently, FRS’s are much richer and more expressive and, hence, much more informative than any other scale based on an unavoidably finite natural language or on its numerical/fuzzy-valued encoding. At this point FRS’s seem to be especially expedient to design many items in questionnaires related to the measurement in Psychology and social sciences.

Unfortunately, and probably because of not being yet enough known, FRS’s have hardly been used and applied, whence the literature about is not yet very extensive (see, for instance, [9, 32, 30, 38, 49, 64, 76, 78, 81]). With this paper we intend to eliminate to a large extent most of the misgivings and qualms about the use of the FRS’s.

2.3. Statistical analysis of fuzzy data: Key notions and features

The need for a methodology to analyze fuzzy data has been frequently highlighted. Hesketh *et al.* [33] pointed out that “... Although further development work is needed, it is possible that fuzzy variables will be able to be used in standard statistical analyses in the traditional manner...” In recent years, this hunch has been proven to be true by considering some appropriate tools, namely, a convenient arithmetic with fuzzy numbers (the one based on Zadeh’s extension principle [94]), a distance between fuzzy numbers (for instance, the one by Diamond and Kloeden [16]), and a suitable mathematical model for the random mechanisms producing fuzzy data, random fuzzy numbers, originally coined as fuzzy random variables in Puri and Ralescu’s sense [63].

In handling data for statistical purposes, the two main algebraic operations to use are the sum and the product by a scalar. In case of fuzzy numbers, the natural extension is the one based on Zadeh’s extension principle [94], which is equivalent to consider level-wise the standard interval arithmetic [57].

More concretely, if \tilde{U} and \tilde{V} are fuzzy numbers, and γ is a real number,

Definition 2.4. The **sum** of \tilde{U} and \tilde{V} is the fuzzy number $\tilde{U} + \tilde{V}$ such that for each $v \in [0, 1]$,

$$(\tilde{U} + \tilde{V})_v = [\inf \tilde{U}_v + \inf \tilde{V}_v, \sup \tilde{U}_v + \sup \tilde{V}_v],$$

and the **product** of \tilde{U} by the scalar γ is the fuzzy number $\gamma \cdot \tilde{U}$ such that for each $v \in [0, 1]$,

$$(\gamma \cdot \tilde{U})_v = \begin{cases} [\gamma \cdot \inf(\tilde{U})_v, \gamma \cdot \sup(\tilde{U})_v] & \text{if } \gamma \geq 0 \\ [\gamma \cdot \sup(\tilde{U})_v, \gamma \cdot \inf(\tilde{U})_v] & \text{if } \gamma < 0 \end{cases}$$

It should be noticed that the space of fuzzy numbers with the preceding operations has not a linear structure, whence the ‘difference between fuzzy numbers’ cannot be properly stated. The loss of linearity of the arithmetic is an important handicap in contrast to what happens for the case of numerical data. So, special care must be taken in the development of statistics with fuzzy data. Actually, the need for an adequate and versatile metric between fuzzy numbers is partially motivated by the need to overcome some of the drawbacks associated with the above mentioned handicap.

A suitable metric for this purpose (although other generalized ones can be found in the literature, e.g., [5]) is the following:

Definition 2.5. *The **2-norm distance** by Diamond and Kloeden [16] between fuzzy numbers \tilde{U} and \tilde{V} is given by*

$$\rho_2(\tilde{U}, \tilde{V}) = \sqrt{\int_{[0,1]} \frac{(\inf \tilde{U}_v - \inf \tilde{V}_v)^2 + (\sup \tilde{U}_v - \sup \tilde{V}_v)^2}{2} dv}.$$

To properly formalize descriptive and inferential statistical data analysis, the random mechanism generating data must be well modelled within the probabilistic setting. In the numerical data case, real-valued random variables model the corresponding mechanism. Following Fréchet [23], who anticipated the idea of abstract random elements taking on values in spaces that are endowed with distances, as well as Féron [20], who adapted Fréchet’s approach to deal with fuzzy values, Puri and Ralescu [63] introduced random fuzzy numbers in their full extent.

Definition 2.6. *Given a random experiment, a mapping \mathcal{X} associating a fuzzy number with each experimental outcome is said to be a **random fuzzy number** (in short **RFN**) (or fuzzy random variable in Puri and Ralescu’s sense) if for all $v \in [0, 1]$ the v -level function is a random interval.*

Equivalently (see, for instance, [27]), \mathcal{X} is an RFN if and only if it is a Borel measurable function with respect to the Borel σ -field generated by ρ_2 . Thanks to this equivalent definition, one can immediately refer to notions like the (induced) *distribution of an RFN*, the *independence of several RFN’s*, and so on, without having to define them expressly. This formalization for the mechanism generating fuzzy data enables RFN’s to preserve/adapt most of the ideas and concepts from the statistical analysis of numerical data (e.g., test p -value, unbiased estimation, statistical consistency of either estimates or tests, etc.) to analyze fuzzy data.

In connection with the analysis of FRS-based items in the context of questionnaires, three crucial features should be recalled, namely:

- There exists an approach involving soundly supported methods and tools to statistically analyze fuzzy data. As it has been pointed out by D’Urso and Gil [17] “... fuzzy data analysis and classification studies are mainly focussed either on developing concepts, results and methods to deal with classical (non-fuzzy) data, where fuzziness is involved in the construction of the analysis/classification procedures (e.g., [7]), or on developing/extending concepts, results and methods concerning data analysis and classification of fuzzy-valued data (e.g., [26]), or on both.” In the context of fuzzy scale-based data, we will essentially follow the second approach.

The main misgiving, which could cause some reluctance to use FRS's, comes probably from the fact that most of the potential practitioners believe fuzzy data cannot be yet properly analyzed from a statistical perspective. In this sense Hesketh *et al.* [30] have underlined that “We are yet to see easily adapted packages that allow for researchers to use the fuzzy concept and then to apply appropriate statistical and other analyses to these in order to both test hypotheses and ensure that meaning is captured.” This aspiration is nowadays a reality that is progressing.

As for the case of numerical data, the induced distribution of an RFN can be summarized by means of several relevant ‘parameters’ or measures, extending relevant ones for random variables. The formal details for these and other parameters can be found in [6], but we are now going to recall the sample version of the two best known ones, which will be used to extend the Cronbach α coefficient.

Definition 2.7. Let $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$ be a sample of fuzzy data from an RFN \mathcal{X} :

- the (Aumann-type) **mean** of $\tilde{\mathbf{x}}$ is given [63] by the fuzzy number $\widetilde{\bar{\mathbf{x}}}$ such that for each $v \in [0, 1]$

$$(\widetilde{\bar{\mathbf{x}}})_v = \left[\frac{\sum_{i=1}^n \inf(\tilde{x}_i)_v}{n}, \frac{\sum_{i=1}^n \sup(\tilde{x}_i)_v}{n} \right];$$

- the (Fréchet-type) **variance** of $\tilde{\mathbf{x}}$ is given by the real number (e.g., [42, 48])

$$s_{\tilde{\mathbf{x}}}^2 = \frac{\sum_{i=1}^n [\rho_2(\tilde{x}_i, \widetilde{\bar{\mathbf{x}}})]^2}{n}.$$

Attempting to develop a methodology for inferential fuzzy data analysis, one should treat each fuzzy datum as a complex unit. Treating their (usually four) characterizing components either separately or as independent ones is neither probabilistically fair nor reliable. By considering a solid mathematical apparatus, it is possible to extend/adapt/develop many concepts, results, ideas, approaches and procedures from Inferential Statistics with numerical data. Along the last years a methodology is being developed to statistically analyze fuzzy rating scale-based data (see [6] for a review), and the involved computations become especially simple when fuzzy data are trapezoidal (see [50]).

It should be pointed out that, on one hand, there are sound mathematical fundamentals within the probabilistic framework (Hilbert spaces,

Borel measurability, etc.) supporting the suitability of the methodology (see, for instance, [6, 27]). On the other hand, as for the numerical data case, the mathematical complexity behind scarcely affects the ease of application and computation of the methodology. Certainly, this methodology to deal with fuzzy data analysis would become more popular if practitioners were aware about the already existing R packages implementing many of the already developed techniques for the analysis of fuzzy data (like those in [46, 82]).

- The shape of fuzzy data has been shown (through extensive simulation studies and real-life examples) to scarcely influence the statistical conclusions from fuzzy data analysis (see [53, 45]). As a consequence, the assumption that arms are linear (i.e., fuzzy responses are trapezoidal) does not entail loss of generality. This unquestionably eases the understanding, implementation and training of the procedure to a great extent, since respondents would need to simply specify the core and the support of their ratings.
- The rating scale involved in designing items in a questionnaire can affect the statistical conclusions. In social research, conclusions are relevant for people, institutions, companies, and policies, so it would be certainly convenient to choose the appropriate scale for measurement. In some simulation-based and real-life examples (see, for instance, [26, 47, 50, 51, 52]) the p -values for tests about means and variances lead to clear significant differences for the FRS but not for the others and conversely.

3. A ‘ranking’ tool between rating scales: The extended α

As it has been already remarked, fuzzy rating scales seem intuitively to be statistically more informative than Likert, VAS and FLS ones. However, no comparison between these scales has been made yet through a unique/unified ‘ranking’ tool, so that these scales could be ranked with respect to a certain criterion. Inspired by researches like the one by Sung and Wu [77], a first convenient tool to start with such a comparison is the one associated with the internal consistency/reliability in the setting of questionnaires.

Constructs in questionnaires are usually approached by means of several statements concerning a focal variable, some of them being quite similar to each other in order to gather enough evidence of the responses in connection with such a variable. The higher the coincidence between values associated with the answers to similar questions or statements the more confident we are

in the individuals having responded in a coherent and comprehensive way. This means a kind of internal consistency, which is a necessary condition for a questionnaire to be useful in terms of reliability.

Cronbach's α coefficient [13] (see, for instance, [11] for a recent history about α and its different versions) has received a great consensus as the most used index for calculating reliability based on the internal consistency. Thus, α is based upon the number of items considered and it is defined as the ratio of true score variance to observed score variance (i.e., as the proportion of the variance of the scale that can be attributed to a common source). Even more, single item measures are mostly seen as unreliable [61]. Cronbach's reliability coefficient is below 1, the closer to 1 the greater the internal consistency of the scale. Actually, it is often considered that values over .9 are excellent and rather needed for individual decision making, and values over .8 are good and acceptable for research purposes (e.g., [60]). Coefficient α is the most frequently considered tool to conclude whether or not items related to the same focal variable elicit consistent and reliable responses, so that if items were replaced by 'similar' ones responses scarcely vary. As it has been recently pointed out by Raykov and Marcoulides [65], in spite of some criticisms, α should remain in service and it actually does.

Since the Fréchet-type variance of fuzzy-valued sample data preserves the essential properties of the real-valued case, one can investigate how to extend Cronbach's α to measure the reliability of focal variables with fuzzy scale-based items. Moreover, one can also develop a comparative investigation among different scales.

3.1. Extension of the Cronbach α for fuzzy data

As anticipated by Hesketh *et al.* [33], the development of work in connection with random fuzzy-valued magnitudes "... means that accepted psychometric standards of validity and reliability can be used to evaluate the potential of this application...". In this respect, and on the basis of the Fréchet-type variance for RFN's, α coefficient [13] could be straightforwardly extended to deal with fuzzy-valued responses as follows:

Definition 3.1. *Given a variable for which a questionnaire involves K related FRS-based items, its **extended Cronbach α** is given by*

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_{j=1}^K s_j^2}{s_{\text{total}}^2} \right),$$

where s_j^2 is the variance of the FLS/FRS-based responses to item j , and s_{total}^2 is the variance of the FLS/FRS-based responses to the sum of all the involved items.

Remark 3.1. Since the studies to be developed in this paper are mostly centered on comparative analyses, a key question to be examined is whether the considered measurement scales are in fact directly comparable or some added scaling should be involved. The answer to this question in connection with the extended Cronbach α is trivial. Thus, since the Fréchet-type variance is squared equivariant under scaling factor and invariant under fuzzy translation, the extended α is invariant under affine transformations. This means that α values, and related conclusions from them, are absolutely irrespective of the bounded reference interval $[l_0, u_0]$ that is chosen for the fuzzy rating scale, so no fuzzy scaling is required.

In the particular case in which the FLS/FRS-based responses to items of the focal variable are trapezoidal, so that $\tilde{x}_{ij} = \text{Tra}(a_{ij}, b_{ij}, c_{ij}, d_{ij})$ is the fuzzy response of the i -th respondent to the j -th item ($i = 1, \dots, n, j = 1, \dots, K$), then the computation of α becomes very simple. More concretely,

$$s_j^2 = \sum_{i=1}^n [\underline{m}_{ij}^2 + \overline{m}_{ij}^2 + \underline{m}_{ij} \cdot \overline{m}_{ij} + \underline{s}_{ij}^2 + \overline{s}_{ij}^2 + \underline{s}_{ij} \cdot \overline{s}_{ij}] / 3n,$$

$$s_{\text{total}}^2 = \sum_{j=1}^K \sum_{i=1}^n [\underline{m}_{ij}^2 + \overline{m}_{ij}^2 + \underline{m}_{ij} \cdot \overline{m}_{ij} + \underline{s}_{ij}^2 + \overline{s}_{ij}^2 + \underline{s}_{ij} \cdot \overline{s}_{ij}] / 3nK,$$

where

$$\begin{aligned} \underline{m}_{ij} &= [(a_{ij} + d_{ij}) - (\overline{a_j} + \overline{d_j})] / 2, & \overline{m}_{ij} &= [(b_{ij} + c_{ij}) - (\overline{b_j} + \overline{c_j})] / 2, \\ \underline{m}_{\cdot j} &= [(\overline{a_j} + \overline{d_j}) - (\overline{a} + \overline{d})] / 2, & \overline{m}_{\cdot j} &= [(\overline{b_j} + \overline{c_j}) - (\overline{b} + \overline{c})] / 2, \\ \underline{s}_{ij} &= [(d_{ij} - a_{ij}) - (\overline{d_j} - \overline{a_j})] / 2, & \overline{s}_{ij} &= [(c_{ij} - b_{ij}) - (\overline{c_j} - \overline{b_j})] / 2, \\ \underline{s}_{\cdot j} &= [(\overline{d_j} - \overline{a_j}) - (\overline{d} - \overline{a})] / 2, & \overline{s}_{\cdot j} &= [(\overline{c_j} - \overline{b_j}) - (\overline{c} - \overline{b})] / 2, \end{aligned}$$

with

$$\begin{aligned} \overline{a_j} &= \sum_{i=1}^{n_j} a_{ij} / n_j, & \overline{b_j} &= \sum_{i=1}^{n_j} b_{ij} / n_j, & \overline{c_j} &= \sum_{i=1}^{n_j} c_{ij} / n_j, & \overline{d_j} &= \sum_{i=1}^{n_j} d_{ij} / n_j. \\ \overline{a} &= \sum_{j=1}^k n_j \cdot \overline{a_j} / n, & \overline{b} &= \sum_{j=1}^k n_j \cdot \overline{b_j} / n, & \overline{c} &= \sum_{j=1}^k n_j \cdot \overline{c_j} / n, & \overline{d} &= \sum_{j=1}^k n_j \cdot \overline{d_j} / n, \\ \underline{m}_{ij} &= \underline{m}_{ij} + \underline{m}_{\cdot j}, & \overline{m}_{ij} &= \overline{m}_{ij} + \overline{m}_{\cdot j}, & \underline{s}_{ij} &= \underline{s}_{ij} + \underline{s}_{\cdot j}, & \overline{s}_{ij} &= \overline{s}_{ij} + \overline{s}_{\cdot j}. \end{aligned}$$

In connection with this extended tool, the following research questions are to be investigated:

Research Question 1: Is the value of the extended α scarcely influenced by the shape of fuzzy responses? Should this be the case, this would help both designing and completing questionnaires involving FRS-based items.

Research Question 2: Is the extended Cronbach's α for FRS-based data behaving analogously to the one for numerical data in connection with the number of items? Should this be the case, this would help to make decisions on the choice of the appropriate number of FRS-based items.

Research Question 3: In applying the extended α to rank different rating scales, can conclusions about majority trends of α be drawn from the comparative analysis? Should this be the case, this would help to conclude which is the most convenient rating scale to consider.

In looking for detailed sound answers to Research Questions 1-3, general conclusions cannot be drawn, whence one cannot state theoretical results. Furthermore, there are not yet suitable realistic models for the distribution of RFN's. However, by examining synthetic examples involving simulations of 'realistic' FRS-based data one can show clear majority trends. The simulation process is partially based on previous ones (see, for instance, [15]), but in analyzing internal consistency some novel assumptions should be added. This makes the whole simulation process a novel crucial endeavor.

3.2. Simulation of FRS-based data in connection with the extended α

To generate fuzzy data from an LU -valued random fuzzy number $\mathcal{X} = LU(a_{\mathcal{X}}, b_{\mathcal{X}}, c_{\mathcal{X}}, d_{\mathcal{X}})$, [70] suggests to use an alternative characterization

$$\mathcal{X} = LU\langle X_1, X_2, X_3, X_4 \rangle,$$

where

$$X_1 = (b_{\mathcal{X}} + c_{\mathcal{X}})/2, \quad X_2 = (c_{\mathcal{X}} - b_{\mathcal{X}})/2, \quad X_3 = b_{\mathcal{X}} - a_{\mathcal{X}}, \quad X_4 = d_{\mathcal{X}} - c_{\mathcal{X}},$$

and LU fuzzy numbers are based on the lower-upper parametric representation by Stefanini *et al.* [74]. LU fuzzy data are generated by simulating the four real-valued random variables X_1, X_2, X_3 and X_4 so that the $\mathbb{R} \times [0, \infty) \times [0, \infty) \times [0, \infty)$ -valued random vector (X_1, X_2, X_3, X_4) will provide us with the 4-tuples (x_1, x_2, x_3, x_4) with $x_1/x_2 = \text{center/radius}$ of the core, and $x_3/x_4 = \text{lower/upper latitudes}$ of the fuzzy number (see Figure 3 for an example of a generated 4-tuple, and the associated LU fuzzy datum when $LU = \text{Tra}$).

To each generated 4-tuple (x_1, x_2, x_3, x_4) we associate the LU fuzzy number

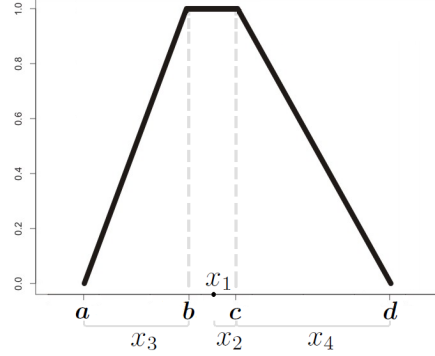


Figure 3: A 4-tuple (x_1, x_2, x_3, x_4) generated from the simulation process, and the associated trapezoidal fuzzy datum

$$LU\langle x_1, x_2, x_3, x_4 \rangle = LU(x_1 - x_2 - x_3, x_1 - x_2, x_1 + x_2, x_1 + x_2 + x_4).$$

For most of the real-life examples one can consider the distribution of X_1 to be skewed. According to the simulation procedure, data have been generated from random fuzzy numbers with a bounded reference set and abstracting and mimicking what has been observed in real-life examples employing the fuzzy rating scale (see [15]). More concretely, fuzzy data have been generated so that

- 5% (or, more generally, $100 \cdot \omega_1\%$) of the data have been obtained by first considering a simulation from a simple random sample of size 4 from a beta $\beta(p, q)$ distribution, the ordered 4-tuple, and finally computing the values of the x_i . The values of p and q vary in most cases to cover six quite different distributions (namely symmetrical weighting central values like $p = q = 1$, $p = q = 2$, symmetrical weighting extreme values like $p = q = 0.75$, and three types of asymmetric ones like $p = 4 > 2 = q$, $p = 6 > 1 = q$ and $p = 6 < q = 10$). The values from the beta distribution should be re-scaled and translated to the reference interval $[l_0, u_0]$.
- 35% (or, more generally, $100 \cdot \omega_2\%$) of the data have been obtained considering a simulation of four random variables $X_i = (u_0 - l_0) \cdot Y_i + l_0$ as follows:

$$\begin{aligned} Y_1 &\sim \beta(p, q), \\ Y_2 &\sim \text{Uniform}[0, \min\{1/10, Y_1, 1 - Y_1\}], \\ Y_3 &\sim \text{Uniform}[0, \min\{1/5, Y_1 - Y_2\}], \\ Y_4 &\sim \text{Uniform}[0, \min\{1/5, 1 - Y_1 - Y_2\}]. \end{aligned}$$

- 60% (or, more generally, $100 \cdot \omega_3\%$) of the data have been obtained considering a simulation of four random variables $X_i = (u_0 - l_0) \cdot Y_i + l_0$ as follows:

$$\begin{aligned}
Y_1 &\sim \beta(p, q), \\
Y_2 &\sim \begin{cases} \text{Exp}(200) & \text{if } Y_1 \in [0.25, 0.75] \\ \text{Exp}(100 + 4 Y_1) & \text{if } Y_1 < 0.25 \\ \text{Exp}(500 - 4 Y_1) & \text{otherwise} \end{cases} \\
Y_3 &\sim \begin{cases} \gamma(4, 100) & \text{if } Y_1 - Y_2 \geq 0.25 \\ \gamma(4, 100 + 4 Y_1) & \text{otherwise} \end{cases} \\
Y_4 &\sim \begin{cases} \gamma(4, 100) & \text{if } Y_1 + Y_2 \geq 0.25 \\ \gamma(4, 500 - 4 Y_1) & \text{otherwise.} \end{cases}
\end{aligned}$$

To adapt this simulation procedure to examine the internal consistency of a construct, novel features and assumptions will be taken into account. In this way, a large sample of $n = 500$ FRS-type data for each of a large number of items, $K = 100$, is to be simulated in accordance with the above described generation procedure. This process will provide with an ‘auxiliary sample’ from which we will later select data for other choices of n and K and transform them **to mimic a certain (linear) dependence**. To generate the 500×100 data we proceed as follows:

- Step 1.** A sample of 500 FRS-type data $(\tilde{x}_1^*, \dots, \tilde{x}_{500}^*)$, the reference interval of the FRS being $[0, 100]$, are first simulated as the ‘auxiliary sample’.
- Step 2.** To mimic the desirable high correlation between the responses from a respondent to different 100 items, for any item j ($j = 1, \dots, 100$)
- a pair (γ_j, δ_j) is considered so that γ_j is generated at random from a uniform distribution in $[0, 1]$ and δ_j is generated from a standard normal distribution;
 - the response of the i -th respondent ($i = 1, \dots, 500$) to the j -th item is assumed to be given by $\tilde{x}_{ij} = \gamma_j \cdot \tilde{x}_i^* + \delta_j + \varepsilon_{ij}$, with ε_{ij} being generated at random from a standard normal distribution;
 - in case any \tilde{x}_{ij} is not fully included within interval $[0, 100]$, the response is appropriately truncated.

Once we get the simulated largest data set including 500×100 fuzzy data, we choose at random and stepwise $n = 450$ from the former 500

respondents, $n = 400$ from the preceding selected 450 respondents, and so on. Analogously, we choose at random and stepwise $K = 50$ from the former 100 items, $K = 40$ from the preceding selected 50 items, and so on. Aiming to be realistic, in some of the studies in the paper, we will constrain K to take on values up to 30.

4. Synthetic STUDY 1: Influence of the shape of fuzzy data on the internal consistency of FRS-based constructs

Regarding the analysis of the effect of the shape of fuzzy data on the value of Cronbach's α , we will consider the above referred LU fuzzy numbers, which could be characterized by means of an ordered 4-tuple (a, b, c, d) , trapezoidal data being an example of them.

This analysis aims to discuss whether or not choosing trapezoidal numbers to model fuzzy scores leads to different α values from those considering some other usual choices.

Certainly, the lack of realistic models for the distributions of random fuzzy numbers prevents us to establish general conclusions. Nevertheless, rather broad conclusions can be drawn by considering empirical examples and, especially, by developing comparative simulation studies between trapezoidal and other considered types of data. In connection with mean and variances, the influence of the shape has been shown to be quite low by inferentially analyzing real-life and simulation examples (e.g., [45, 52]).

In connection with the extended α , the influence of the shape is to be descriptively analyzed. The types of fuzzy numbers involved in this subsection attempt to be close representations of rather similar imprecise scores, and they correspond (see Figure 4) to

- six of the LU 's (e.g., [73, 74]) that can be characterized by means of four real numbers (namely, the extremes of their core and support); in particular, we have chosen trapezoidal fuzzy numbers (Tra) *vs* quadratic functions (Π -curves), functions with parametric monotonic interpolation either using rational splines (LU_{1A} and LU_{1B}), or mixed exponential splines (LU_{2A} and LU_{2B});
- two additional fuzzy numbers have been added to the comparison, namely, triangular $\text{Tri}(a, b, c, d) = \text{Tra}(a, (b + c)/2, (b + c)/2, d)$, and symmetric triangular $\text{TriS}(a, b, c, d) = \text{Tra}(a, (a + d)/2, (a + d)/2, d)$.

Table 1 collects the extended Cronbach's α values for a construct based on several samples of FRS-based simulated data. This table only shows choices of K up to 30, since outputs for larger values are similar.

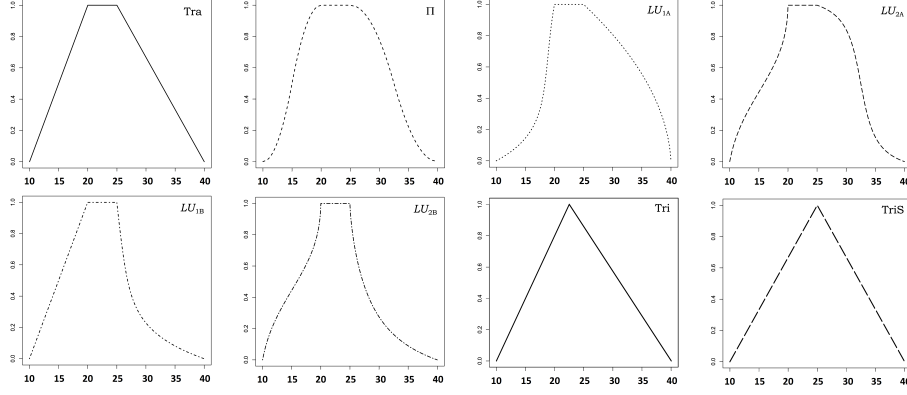


Figure 4: Eight types of fuzzy numbers sharing support, $(10, 40)$, the first six ones sharing also the core, $[20, 25]$, and all of them differing in shape

Table 1: Cronbach's α for a construct in simulated samples, in accordance with different considered shapes for the FRS-based responses

n & K choices / shape		Tra	II	LU_{1A}	LU_{1B}	LU_{2A}	LU_{2B}	Tri	TriS
$n = 300$	$K = 30$	0.9031	0.9029	0.9025	0.9039	0.9030	0.9041	0.9024	0.8999
	$K = 25$	0.8763	0.8761	0.8756	0.8773	0.8762	0.8776	0.8753	0.8722
	$K = 20$	0.8678	0.8676	0.8672	0.8687	0.8677	0.8690	0.8668	0.8638
	$K = 15$	0.8281	0.8279	0.8274	0.8293	0.8280	0.8296	0.8270	0.8234
	$K = 10$	0.8106	0.8103	0.8099	0.8120	0.8106	0.8124	0.8096	0.8054
	$K = 7$	0.7655	0.7652	0.7646	0.7670	0.7655	0.7674	0.7641	0.7596
$n = 100$	$K = 5$	0.6255	0.6251	0.6238	0.6274	0.6253	0.6279	0.6240	0.6190
	$K = 30$	0.9114	0.9114	0.9110	0.9124	0.9114	0.9127	0.9109	0.9081
	$K = 25$	0.8851	0.8850	0.8845	0.8865	0.8851	0.8868	0.8844	0.8806
	$K = 20$	0.8775	0.8774	0.8769	0.8789	0.8775	0.8792	0.8769	0.8732
	$K = 15$	0.8451	0.8450	0.8442	0.8468	0.8450	0.8472	0.8444	0.8404
	$K = 10$	0.8241	0.8240	0.8229	0.8262	0.8240	0.8266	0.8235	0.8189
$n = 50$	$K = 7$	0.7891	0.7890	0.7877	0.7916	0.7890	0.7922	0.7885	0.7827
	$K = 5$	0.6276	0.6275	0.6244	0.6320	0.6274	0.6328	0.6275	0.6196
	$K = 30$	0.9267	0.9267	0.9262	0.9276	0.9267	0.9278	0.9266	0.9243
	$K = 25$	0.9027	0.9027	0.9019	0.9041	0.9027	0.9044	0.9026	0.8994
	$K = 20$	0.8963	0.8963	0.8957	0.8976	0.8963	0.8979	0.8962	0.8930
	$K = 15$	0.8578	0.8577	0.8569	0.8596	0.8577	0.8599	0.8577	0.8535
$n = 30$	$K = 10$	0.8278	0.8276	0.8265	0.8302	0.8277	0.8307	0.8275	0.8218
	$K = 7$	0.7875	0.7873	0.7854	0.7911	0.7873	0.7918	0.7870	0.7789
	$K = 5$	0.6071	0.6068	0.6047	0.6126	0.6069	0.6138	0.6063	0.5924
	$K = 30$	0.9171	0.9170	0.9160	0.9184	0.9170	0.9187	0.9169	0.9145
	$K = 25$	0.8897	0.8896	0.8880	0.8915	0.8895	0.8918	0.8895	0.8865
	$K = 20$	0.8895	0.8895	0.8882	0.8911	0.8894	0.8913	0.8894	0.8868
$n = 10$	$K = 15$	0.8463	0.8462	0.8443	0.8485	0.8461	0.8489	0.8460	0.8423
	$K = 10$	0.8192	0.8191	0.8168	0.8220	0.8189	0.8224	0.8188	0.8142
	$K = 7$	0.7842	0.7840	0.7811	0.7879	0.7839	0.7885	0.7836	0.7772
	$K = 5$	0.5977	0.5974	0.5937	0.6032	0.5973	0.6042	0.5967	0.5866
	$K = 30$	0.9151	0.9150	0.9119	0.9172	0.9147	0.9174	0.9148	0.9137
	$K = 25$	0.8893	0.8892	0.8851	0.8919	0.8889	0.8921	0.8890	0.8883
$n = 10$	$K = 20$	0.8884	0.8883	0.8846	0.8908	0.8880	0.8910	0.8883	0.8875
	$K = 15$	0.8263	0.8261	0.8206	0.8299	0.8257	0.8302	0.8255	0.8238
	$K = 10$	0.8026	0.8025	0.7957	0.8065	0.8019	0.8067	0.8016	0.8009
	$K = 7$	0.8104	0.8103	0.8043	0.8138	0.8098	0.8140	0.8100	0.8106
	$K = 5$	0.5329	0.5328	0.5167	0.5404	0.5314	0.5403	0.5302	0.5373

On the basis of these simulation-based results one can conclude that the chosen shape scarcely affects the internal reliability of the constructs. For this reason, in the following studies we only consider trapezoidal data.

5. Synthetic STUDY 2: Effect of the number of items on the internal consistency of FRS-based constructs

In this study, and given the scarce influence of the shape of fuzzy-valued responses, fuzzy data are generated in accordance with the method in the Synthetic STUDY 1 by considering $LU = \text{Tra}$. This analysis aims to examine the effect of the number of items in a construct, and secondarily that of the number of respondents, on the value of the extended α . Such an influence has been widely reported in the literature in connection with Likert-type scales (e.g., [35, 36]).

Different choices of K and n have been considered, and the corresponding values of α are displayed in tabular and graphical ways. First, values of the extended α for all the considered choices of the number of items and several sample sizes are collected in Table 2, and later a slightly more detailed information is gathered in Figure 5.

Table 2: Cronbach's α for the simulated data and different choices of n (number of respondents) and K (number of items)

K / n	500	400	300	200	150	100	50	40	30	10
100	0.9660	0.9659	0.9657	0.9655	0.9683	0.9686	0.9717	0.9701	0.9683	0.9659
50	0.9356	0.9354	0.9352	0.9340	0.9392	0.9398	0.9482	0.9449	0.9438	0.9437
40	0.9197	0.9196	0.9194	0.9193	0.9260	0.9262	0.9359	0.9314	0.9304	0.9270
30	0.9036	0.9036	0.9031	0.9014	0.9096	0.9114	0.9267	0.9208	0.9171	0.9151
25	0.8769	0.8772	0.8763	0.8732	0.8839	0.8851	0.9027	0.8942	0.8897	0.8893
20	0.8650	0.8667	0.8678	0.8653	0.8771	0.8775	0.8963	0.8867	0.8895	0.8884
15	0.8210	0.8245	0.8281	0.8293	0.8459	0.8451	0.8578	0.8380	0.8463	0.8263
10	0.8008	0.8077	0.8106	0.8155	0.8324	0.8241	0.8278	0.8129	0.8192	0.8026
7	0.7585	0.7631	0.7655	0.7699	0.7943	0.7891	0.7875	0.7794	0.7842	0.8104
5	0.6203	0.6281	0.6255	0.6374	0.6695	0.6276	0.6071	0.5862	0.5977	0.5329

On the basis of the outputs in Table 2 and Figure 5, the influence of the number of items on the value of the extended α is unequivocally clear. Additionally, by looking at Table 2 we can check that the effect of the number of respondents on the value of the extended α is not so relevant (see [3, 8, 36, 91] for some related studies in case of Likert scales).

6. Synthetic STUDY 3: Influence of the rating scale on the internal consistency of constructs

In developing comparative studies in connection with **Research Question 3** there is a need, on one hand, for processes to encode Likert responses both numerically and linguistically fuzzy. On the other hand, because we have to consider simulation studies, there is a need to mimic the link between FRS and Likert responses, as well as the one between FRS and VAS responses.

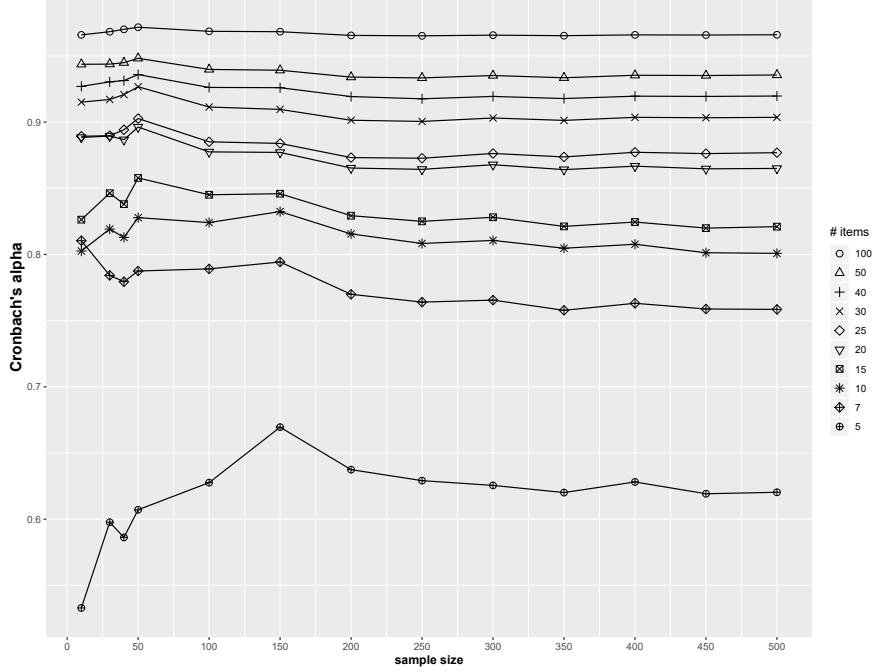


Figure 5: Evolution of Cronbach's α with respect to the sample size for different numbers of items in the construct

As for the Synthetic STUDY 1, in the simulations in this study the reference interval will be assumed to be $[l_0, u_0] = [0, 100]$. Furthermore, in the comparative developments between scales, the Likert scale will be considered to be a 5-point one, $\{L_1, L_2, L_3, L_4, L_5\}$.

6.1. Mimicking the link between rating scales

To make the comparison between rating scales being really fair, the **numerically encoded Likert** NEL that will be considered is the usual one (1, 2, 3, 4, and 5), but rescaled in accordance with the reference interval, so that $L_1 \equiv 0$, $L_2 \equiv 25$, $L_3 \equiv 50$, $L_4 \equiv 75$, $L_5 \equiv 100$. For a k -point scale with reference interval $[l_0, u_0]$, the numerical encoding would be to rescale i to be $L_i \equiv L_i = l_0 + (u_0 - l_0) \cdot (i - 1) / (k - 1)$, for $i \in \{1, \dots, k\}$.

The **fuzzy linguistic encodings** will be chosen among some of the most frequently used fuzzy linguistic scales when 5 options are involved (see [29, 58, 89]). The four chosen fuzzy linguistic encodings (fulfilling the so-called Ruspini condition [67]) have been displayed in Figure 6, FLS_3^5 being the balanced semantic representation of the 5 linguistic hierarchies, and FLS_4^5

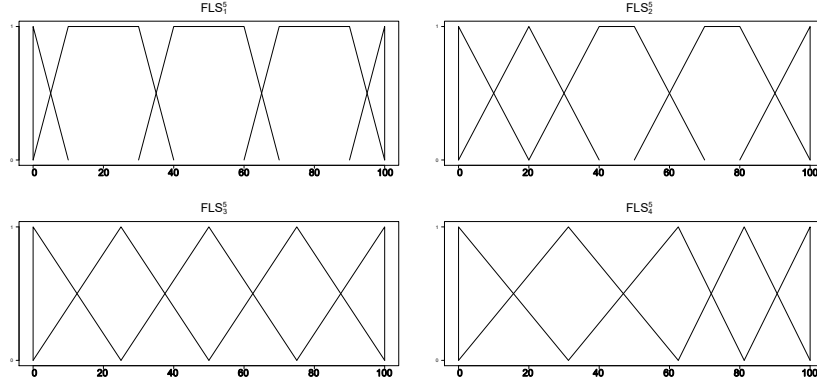


Figure 6: Frequently used fuzzy linguistic encodings of 5-point Likert scales

being an asymmetric representation inspired by the unbalanced semantic of the same number of labels [29].

In some previous real-life examples involving questionnaires (see, for instance, [27, 47, 50, 52]), several items were designed by considering a double type scale (Likert+FRS) to respond questions, and components of each double response are linked, since they come from the same respondent. However, this link cannot be immediately stated and should be mimicked in simulation processes. A reasonable **‘Likertization’ process** can be considered in terms of the numerical encoding and based on the *minimum distance criterion*, and later a fuzzy linguistic encoding can be applied.

Definition 6.1. *If the considered Likert scale is a k -point one and the reference interval for the FRS is $[l_0, u_0]$, the minimum distance criterion consists in associating each FRS-based datum with the number in $NEL = \{L_1, \dots, L_k\}$ with the smallest distance to the given datum. That is, each fuzzy datum \tilde{U} is associated with the numerical value $\kappa(\tilde{U})$ such that*

$$\kappa(\tilde{U}) = \arg \min_{L_i, i \in \{1, \dots, k\}} \rho_2(\tilde{U}, \mathbb{1}_{\{L_i\}}),$$

$\mathbb{1}_A$ denoting the indicator function of set A . In particular,

$$\kappa(\text{Tra}(a, b, c, d)) = \arg \min_{L_i, i \in \{1, \dots, k\}} [2 \cdot L_i^2 - (a + b + c + d) \cdot L_i],$$

which corresponds to the L_i being closer to the ‘central point’ $(a+b+c+d)/4$.

If in solving the above minimization we get two involved coincident distances, the associated numerically encoded Likert response can be chosen at random among the two corresponding numerical values. It should be pointed out that other reasonable and well-supported Likertization criteria

(like, for instance, the supervised classification in [12]) have been considered, but they are much more complex and almost generally lead to a very close validation.

In a similar way, if one wishes to mimic a double type scale (VAS+FRS) to respond questions, and components of each double response are assumed to be linked, since they come from the same respondent, a defuzzification procedure could be considered. A reasonable ‘**defuzzification**’ process is the one introduced in [88], and later extended (see [59, 68]) as the *weighted average based on levels*.

Definition 6.2. *The weighted averaging based on levels (in short WABL) of fuzzy number \tilde{U} is the real number $\text{WABL}(\tilde{U})$ such that*

$$\text{WABL}(\tilde{U}) = \int_{[0,1]} \frac{\inf \tilde{U}_v + \sup \tilde{U}_v}{2} dv.$$

It can be easily proved that WABL is equivariant under linear transformations. Moreover, for trapezoidal fuzzy numbers we have that

$$\text{WABL}(\text{Tra}(a, b, c, d)) = (a + b + c + d)/4,$$

that is, the ‘central point’ which is the mid-point of the core and the support ones and, hence, equivalent to the mid-point of the .5-level.

6.2. Analysis

This analysis aims to examine whether or not the choice of the rating scale to respond questions influences the value of the extended Cronbach α . Although a general conclusion about cannot be drawn, one can easily state a very majority trend. For different choices of n (number of respondents) and K (number of items), and by following the scheme leading to data in Table 2, 1000 samples of $n \times K$ FRS-based data have been generated, and later encoded by means of the already described processes. Later, percentages of samples for which Cronbach’s α of the FRS-based data is greater than that of the VAS-defuzzified FRS-based and the encoded Likert ones are computed and collected. Table 3 shows a few choices of K up to 30, albeit outputs for larger values are similar.

Consequently, in getting a larger internal consistency of a construct, majority trends support the almost general superiority of the FRS with respect to the VAS, the high superiority of the VAS with respect to the NEL and the superiority of this one with respect to the FLS’s.

Table 3: Percentages of simulated samples for which Cronbach’s α of some rating scales are greater than that of other ones for different choices of n (number of respondents) and K (number of items)

n & K choices / scales		$\alpha_{\text{FRS}} > \alpha_{\text{VAS}}$	$\alpha_{\text{VAS}} > \alpha_{\text{NEL}}$	$\alpha_{\text{NEL}} > \alpha_{\text{FLS}_1^5}$	$\alpha_{\text{NEL}} > \alpha_{\text{FLS}_2^5}$	$\alpha_{\text{NEL}} > \alpha_{\text{FLS}_3^5}$	$\alpha_{\text{NEL}} > \alpha_{\text{FLS}_4^5}$
$n=300$	$K=30$	100	100	100	100	100	100
	$K=20$	100	100	99.9	99.9	99.9	100
	$K=10$	100	99.1	99.5	97.7	97.9	98.4
	$K=5$	99.2	94.0	92.7	90.0	91.6	92.5
$n=100$	$K=30$	100	99.9	100	99.4	99.6	99.8
	$K=20$	100	99.8	98.8	97.3	98.0	98.4
	$K=10$	99.9	96.0	94.1	90.3	91.5	92.2
	$K=5$	97.7	84.6	84.7	82.2	82.1	81.6
$n=50$	$K=30$	100	99.5	98.3	96.4	97.8	98
	$K=20$	99.9	98.0	95.9	92.7	94.3	94.6
	$K=10$	99.3	89.79	87.3	83.4	86.2	85.0
	$K=5$	94.9	76.6	77.8	76.1	75.3	73.9
$n=30$	$K=30$	99.7	98.3	95.8	91.4	93.4	94.4
	$K=20$	99.6	95.4	91.6	86.6	89.5	90.3
	$K=10$	97.5	85.5	82.5	78.4	79.8	79.7
	$K=5$	92.5	73.2	71.3	69.5	69.3	69.4

7. Empirical STUDY 4: illustrating with a real-life example

In this study, the majority trends in the synthetic ones are to be illustrated by means of a real-life example.

A **sample** of 70 people (from Spain and Italy) with different age, background, work type and position has been considered to respond to a questionnaire about *Food Quality And Satisfaction Evaluation* that measures two constructs, namely, the ‘quality of the food and beverage’ (FB) and the ‘satisfaction with restaurant services’ (RS). The items have consisted of making the customers answer to some numerical or dichotomic (YES/NO) questions, and also to rate the degree of agreement with several assertions in connection with the restaurant they visit more often to have lunch. Questions have been mostly selected among the 20 in <https://www.questionpro.com/survey-templates/fast-food-restaurant/>.

The selected questions that are related to a rather imprecise answer have been those associated with FB

- FB1: The food is served hot and fresh
- FB2: The menu has a good variety of items
- FB3: The quality of food is excellent
- FB4: The food is tasty and flavorful
- FB5: The quality of beverage is good,

and those associated with RS

- RS1: My food order was correct and complete
- RS2: Employees are patient when taking my order
- RS3: I was served promptly

- RS4: Good availability of sauces, utensils, napkins,...
- RS5: The menu board was easy to read
- RS6: Employees are friendly and courteous
- RS7: The service is excellent
- RS8: Good cleanness of the restaurant and service.

Concerning these 13 questions a double rating scale, namely, the original 5-point Likert scale where pre-specified responses have been SD = ‘STRONGLY DISAGREE’, sd = ‘SOMEWHAT DISAGREE’, n = ‘NEUTRAL’, sa = ‘SOMEWHAT AGREE’, and SA = ‘STRONGLY AGREE’, and an FRS with referential interval $[0,100]$ and trapezoidal answers.

Figure 7 displays an excerpt of the questionnaire form in connection with one of the double-rating items. The precise questions along with the ones related to focal variables FB and RS can be found detailed in <http://bellman.ciencias.uniovi.es/srabss/Archivos/Restaurants-SupplementaryMaterial.pdf>.

Figure 7: Excerpt of the questionnaire form in connection with the double-rating item FB3: The quality of food is excellent

The **training** to respond Items FB1-FB5 and RS1-RS8 have been mostly carried out by Master’s students in a course on Fuzzy Statistics. Respondents were selected to integrate a convenient sample, their ages, backgrounds, working status, etc. being diverse. The training has been possible in a moderate time (always up to fifteen minutes).

The FRS-based **data** have been displayed in gray, those corresponding to responses to FB1-FB5 in Figure 8, and those corresponding to RS1-RS8 in Figure 9. In each of the thirteen items, the associated sample Aumann-type mean has been also displayed in black.

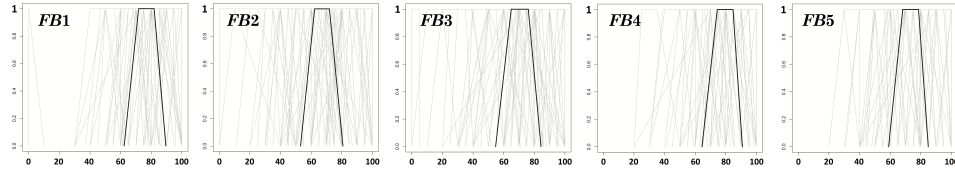


Figure 8: From left to right FRS-based responses to Questions FB1 to FB5 (in gray) along with their sample mean responses (in black bold) for the real-life example

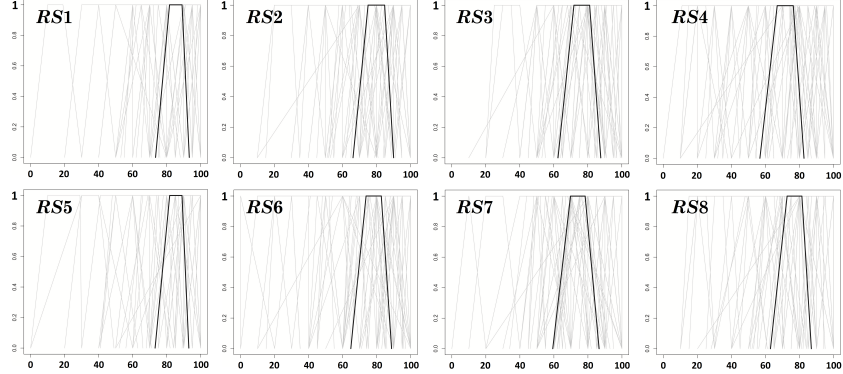


Figure 9: From left to right, and top to bottom, FRS-based responses to Questions RS1 to RS8 (in gray) along with their sample mean responses (in black bold) for the real-life example

The analysis in this empirical study corroborates results in Synthetic STUDIES 1 to 3 in terms of the real-life dataset. For STUDY 3, the encodings of the Likert responses and the defuzzification of the FRS-based ones have followed the processes described in such a study. Constructs have been $FB \equiv$ ‘quality of the food and beverage’, $RS \equiv$ ‘satisfaction with restaurant services’, and $Global \equiv$ ‘overall satisfaction with restaurant’.

Table 4: Cronbach’s α for the real-life example in accordance with different considered shapes for the FRS-based responses

shape \ construct	FB (5 items)	RS (8 items)	Global (13 items)
Tra	0.6928	0.8068	0.8472
II	0.6921	0.8063	0.8466
LU_{1A}	0.6886	0.8046	0.8457
LU_{1B}	0.6887	0.8087	0.8466
LU_{2A}	0.6929	0.8070	0.8473
LU_{2B}	0.6875	0.8092	0.8466
Tri	0.6857	0.8023	0.8426
TriS	0.6920	0.7924	0.8398

Table 5: Cronbach’s α for the two constructs and global in the real-life example for different rating scales

rating scale \ construct	FB (5 items)	RS (8 items)	Global (13 items)
FRS	0.6928	0.8068	0.8472
VAS	0.6696	0.7934	0.8333
NEL	0.5660	0.7465	0.7775
FLS_1^5	0.5406	0.7459	0.7756
FLS_2^5	0.5502	0.7455	0.7742
FLS_3^5	0.5530	0.7479	0.7778
FLS_4^5	0.5347	0.7531	0.7832

Table 4 shows the values of the extended α for different shapes, which agree with those from the Synthetic STUDY 1. Table 5 shows Cronbach's α for different constructs and rating scales. Results agree with those from the Synthetic STUDY 3.

Both Table 4 and Table 5 also support the conclusions from the Synthetic STUDY 2. Thus, the larger the number of items in a construct the larger its internal consistency.

8. General discussion and concluding remarks

The preceding studies allow us to state several summary assertions concerning the majority behavior of the extended α with respect to: the shape of the fuzzy scores, the number of items in a construct where items are FRS-based, and the considered rating scale.

By looking at the rows in Table 1, the results in connection with **Research Question 1** lead to conclude that the shape modelling fuzzy responses scarcely influences the reliability based on the internal consistency. An implication from this is that trapezoidal responses can be considered, and this will certainly ease designing and filling out such questionnaires.

The results in connection with **Research Question 2** allow us to conclude that for almost all the simulated samples the larger the number of items in a construct the greater the Cronbach α . Actually, for over 10 items, the reliability based on internal consistency is usually over .8. Regarding the number of respondents, the sample size has not necessarily a positive impact on the value of α , but the greater the sample size the greater the accuracy of the estimation of the population reliability. Figure 5 shows another majority trend: constructs involving more than 10 items are quite robust with respect to the sample size. These conclusions can be valuable to help in making a decision about the convenient number of items to guarantee a given α .

Finally, the results in connection with **Research Question 3** lead to state that Cronbach's α is mostly larger for the FRS-based data than for the VAS-based ones, for the latter it is quite mostly larger than for the numerically encoded Likert ones, and it is mostly larger for the numerical than for the fuzzy linguistically encoded Likert scores, as summarized in Figure 10.

Consequently, as long as the conclusions from questionnaires are going to have a big impact on future decisions, policies, and so on, it would be recommended to consider fuzzy rating scales-based items instead of Likert- or visual analogue-type ones. The first ones allow researchers exploiting much more information because of capturing individual differences, and reflecting


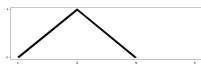
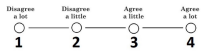

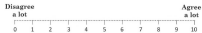

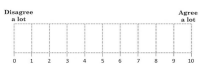
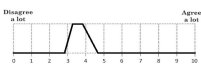
RATING SCALE TYPE	rating scale frame	example of response	captures intrinsic imprecision	captures individual differences	majority trend for α
Fuzzy Linguistically encoded Likert			partially	NO ✗	\wedge
Numerically encoded Likert			NO ✗	NO ✗	\wedge
Visual Analogue			NO ✗	YES ✓	\wedge
Computerized Fuzzy Rating			YES ✓	YES ✓	

Figure 10: Comparative summary of rating scales with respect to α , showing how reliability based on internal consistency increases from top to bottom (\wedge = lower than)

the intrinsic imprecision underlying the psychological measurement, so that the internal consistency of questionnaires mostly benefits from the use of the fuzzy rating scales.

Concerning future directions, on one hand other interesting scales and methods can be found in recent psychological studies. They could be considered in future ones for purposes of analyzing and comparing with the scales in this paper. Among them one can mention the Visual Analogue Scale for Rating, Ranking and Paired-comparison (VA-RRP) by Sung and Wu [77], being this last one an especially appealing scale in the context of this paper.

Despite alternative methods for estimating reliability have been suggested in the literature, like coefficients *omega* and *H*, the greatest lower bound, and so on [10, 54]. Nonetheless, α remains as the first and simplest to compute choice in most occasions in applied and research contexts (for recent instances, see [2, 84]).

Acknowledgements

The authors are very grateful to the reviewers in charge of the manuscript for their valuable comments and suggestions.

References

- [1] G. Acampora, P. Foggia, A. Saggese, M. Vento, A hierarchical neuro-fuzzy architecture for human behavior analysis, *Inf. Sci.* 310 (2015) 130–148.
- [2] Y. Alkis, Z. Kadirhan, M. Sat, Development and validation of social anxiety scale for social media users, *Comput. Hum. Behav.* 72 (2017) 296–303.
- [3] W.N. Arifin, A web-based sample size calculator for reliability studies, *Educ. Med. J.* 10(3) (2018) 67–76.

- [4] E. Benoit, Expression of uncertainty in fuzzy scales based measurements. *Measurement* 46 (2013) 3778–3782.
- [5] C. Bertoluzza, N. Corral, A. Salas, On a new class of distances between fuzzy numbers, *Mathw. Soft Comput.* 2 (1995) 71–84.
- [6] Á. Blanco-Fernández, M.R. Casals, A. Colubi, N. Corral, M. García-Bárzana, M.Á. Gil, G. González-Rodríguez, M.T. López, M.A. Lubiano, M. Montenegro, A.B. Ramos-Guajardo, S. De la Rosa de Saa, B. Sinova, A distance-based statistical analysis of fuzzy number-valued data. *Int. J. Approx. Reasoning*, 55(7) (2014) 1487–1501; Rejoinder, *Int. J. Approx. Reasoning* 55(7) (2014) 1601–1605.
- [7] J.H. Bolin, J.M. Edwards, W.H. Finch, J.C. Cassady, Applications of cluster analysis to the creation of perfectionism profiles: a comparison of two clustering approaches, *Front. Psychol.* 5 (2014) Art. 343.
- [8] D.G. Bonett, T.A. Wright, Cronbach’s alpha reliability. Interval estimation, hypothesis testing, and sample size planning, *J. Organ. Behav.* 36(1) (2014) 3–15.
- [9] A. Castro-López, J.M. Alonso, Modeling human perceptions in e-commerce applications: A case study on business-to-consumers websites in the textile and fashion sector, In: A. Meier, E. Portmann, & L. Terán (Eds.), *Applying Fuzzy Logic for the Digital Economy and Society. Fuzzy Management Methods*, Springer, Cham, 2019, pp. 115–134.
- [10] E. Cho, Making reliability reliable: A systematic approach to reliability coefficients, *Organ. Res. Methods* 19(4) (2016) 651–682.
- [11] E. Cho, S. Kim, Chronbach’s coefficient alpha: Well-known but poorly understood, *Organ. Res. Methods* 18(2) (2015) 207–230.
- [12] A. Colubi, G. González-Rodríguez, M.Á. Gil, W. Trutschnig, Nonparametric criteria for supervised classification of fuzzy data, *Int. J. Approx. Reasoning* 52 (2011) 1272–1282.
- [13] L.J. Cronbach, Coefficient alpha and the internal structure of tests, *Psychometrika* 16(3) (1951) 297–334.
- [14] M. Dağdeviren, I. Yüksel, Developing a fuzzy analytic hierarchy process (AHP) model for behavior-based safety management, *Inf. Sci.* 178 (2008) 1717–1733.
- [15] S. de la Rosa de Saa, M.Á. Gil, G. González-Rodríguez, M.T. López, M.A. Lubiano, Fuzzy rating scale-based questionnaires and their statistical analysis, *IEEE Trans. Fuzzy Syst.* 23(1) (2015) 111–126.
- [16] P. Diamond, P. Kloeden, Metric spaces of fuzzy sets, *Fuzzy Sets Syst.* 35 (1990) 241–249.
- [17] P. D’Urso, M.Á. Gil, Fuzzy data analysis and classification, *Adv. Data Anal. Classif.* 11(4) (2017) 645–657.
- [18] B. Efe, M. Kurt, A systematic approach for an application of personnel selection in assembly line balancing problem, *Int. Trans. Oper. Res.* 25(3) (2018) 1001–1025.
- [19] S.Z. Eyupoglu, L.A. Gardashova, R.A. Allahverdiyev, T. Saner, Application of fuzzy logic in job satisfaction performance problem, *Procedia Computer Science* 102 (2016) 190–197.
- [20] R. Féron, Ensembles aléatoires flous, *Comptes Rendus Acad. Sci. Ser. A* 282 (1976) 903–906.
- [21] P.C. Fiss, Building better causal theories: A fuzzy set approach to typologies in organization research, *Acad. Manage. J.* 54(2) (2011) 393–420.

- [22] M.L. Frazier, C. Tupper, S. Fainshmidt, The path(s) to employee trust in direct supervisor in nascent and established relationships: A fuzzy set analysis, *J. Organ. Behav.* 37 (2016) 1023–1043.
- [23] M. Fréchet, Les éléments aléatoires de nature quelconque dans un espace distancié, *Ann. Inst. Henri Poincaré* 10 (1948) 215–310.
- [24] F. Funke, U.D. Reips, Why semantic differentials in Web-based research should be made from visual analogue scales and not from 5-point scales, *Field Methods* 24 (2012) 310–327.
- [25] A.S. Gabriel, J.T. Campbell, E. Djurdjevic, R.E. Johnson, C.C. Rosen, Fuzzy profiles: Comparing and contrasting Latent Profile Analysis and Fuzzy Set Qualitative Comparative Analysis for person-centered research, *Organ. Res. Methods* 21(4) (2018) 877–904.
- [26] M.Á. Gil, M.A. Lubiano, S. de la Rosa de Sáa, B. Sinova, Analyzing data from a fuzzy rating scale-based questionnaire. A case study, *Psicothema* 27(2) (2015) 182–191.
- [27] G. González-Rodríguez, A. Colubi, M.Á. Gil, Fuzzy data treated as functional data. A one-way ANOVA test approach, *Comput. Stat. Data Anal.* 56(4) (2012) 943–955.
- [28] M.H.S. Hayes, D.G. Patterson, Experimental development of the graphic rating method, *Psychol. Bull.* 18 (1921) 98–99.
- [29] F. Herrera, E. Herrera-Viedma, L. Martínez, A fuzzy linguistic methodology to deal with unbalanced linguistic term sets, *IEEE Trans. Fuzzy Syst.* 16(2) (2008) 354–370.
- [30] B. Hesketh, B. Griffin, V. Loh, A future-oriented retirement transition adjustment framework, *J. Vocat. Behav.* 79(2) (2011) 303–314.
- [31] T. Hesketh, B. Hesketh, Computerized fuzzy ratings: The concept of a fuzzy class, *Behav. Res. Methods Instr. Comput.* 26(3) (1994) 272–277.
- [32] B. Hesketh, K. McLachlan, D. Gardner, Work adjustment theory: An empirical test using a fuzzy rating scale, *J. Vocat. Behav.* 40(3) (1992) 318–337.
- [33] B. Hesketh, R. Pryor, M. Gleitzman, T. Hesketh, Practical applications and psychometric evaluation of a computerised fuzzy graphic rating scale, In: T. Zétényi (Ed.) *Fuzzy Sets in Psychology. Advances in Psychology Series*, 56(C), NH/Elsevier, Amsterdam, 1988, pp. 425–454.
- [34] T. Hesketh, R., Pryor, B. Hesketh, An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences, *Int. J. Man-Machine Stud.* 29(1) (1988) 21–35.
- [35] R. Hoekstra, J. Vugteveen, M.J. Warrens, P.M. Kruijen, An empirical analysis of alleged misunderstandings of coefficient alpha, *Int. J. Soc. Res. Methodol.* 22(4) (2019) 351–364.
- [36] S.B. Javali, N.V. Gudaganavar, S.M. Raj, Effect of varying sample size in estimation of reliability coefficients of internal consistency, *WebmedCentral BIO-STATISTICS* 2(2) (2011) WMC001572.
- [37] K.Y. Jin, H.F. Chen, MIMIC approach to assessing differential item functioning with control of extreme response style, *Behav. Res. Methods*, 52 (2020) 23–35.
- [38] T. Jónás, Z.E. Tóth, G. Árva, Applying a fuzzy questionnaire in a peer review process, *Total Qual. Manag. Bus. Excell.* 29(9–10) (2018) 1228–1245.

- [39] C. Karyotis, F. Doctor, R. Iqbal, A. James, V. Chang, A fuzzy computational model of emotion for cloud based sentiment analysis. *Inf. Sci.* 433–434 (2018) 448–463.
- [40] M. Kochen, Applications of fuzzy sets in Psychology. In: L.A. Zadeh, K.-S. Fu, K. Tanaka, M. Shimura (Eds.), *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, Academic Press, New York, 1975, pp. 395–408.
- [41] J. Konacoglu, I. Albayrak, A new fuzzy decision making approach for personnel selection problem, *Intell. Dec. Tech.* 12(4) (2019) 471–482.
- [42] R. Körner, On the variance of fuzzy random variables, *Fuzzy Sets Syst.* 92(1) (1997) 83–93.
- [43] T. Kuhlmann, M. Dantlgraber, U.D. Reips, Investigating measurement equivalence of visual analogue scales and Likert-type scales in Internet-based personality questionnaires, *Behav. Res. Methods* 49(6) (2017) 2173–2181.
- [44] R. Likert, A technique for the measurement of attitudes, *Archives of Psychology* 22 (1932) 140–155.
- [45] M.A. Lubiano, C. Carleos, M. Montenegro, M.Á. Gil, Case study-based sensitivity analysis of scale estimates w.r.t. the shape of fuzzy data, In: S. Destercke, T. Denoeux, M.Á. Gil, P. Grzegorzewski, & O. Hryniewicz (Eds.), *Uncertainty Modelling in Data Science, Advances in Intelligent Systems and Computing*, 832, Springer, Heidelberg, 2019, pp. 157–165.
- [46] M.A. Lubiano, S. de la Rosa de Saa, *FuzzyStatTra: Statistical Methods for Trapezoidal Fuzzy Numbers* (2017) Version 1.0.
- [47] M.A. Lubiano, S. de la Rosa de Saa, M. Montenegro, B. Sinova, M.Á. Gil, Descriptive analysis of responses to items in questionnaires. Why not using a fuzzy rating scale?, *Inf. Sci.* 360 (2016) 131–148.
- [48] M.A. Lubiano, M.Á. Gil, M. López-Díaz, M.T. López, The $\vec{\lambda}$ -mean squared dispersion associated with a fuzzy random variable, *Fuzzy Sets Syst.* 111 (2000) 307–317.
- [49] M.A. Lubiano, P. González-Gil, H. Sánchez-Pastor, C. Pradas, H. Arnillas, An incipient Fuzzy Logic-based analysis of the medical specialty influence on the perception about mental patients. In: E. Gil, E. Gil, J. Gil, & M. Á. Gil (Eds.) *The Mathematics of the Uncertain, Studies in Systems, Decision and Control*, 142, Springer, Heidelberg, 2018, pp. 653–662.
- [50] M.A. Lubiano, M. Montenegro, B. Sinova, S. de la Rosa de Saa, M.Á. Gil, Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications, *Eur. J. Oper. Res.* 251(3) (2016) 918–929.
- [51] M.A. Lubiano, A. Salas, C. Carleos, S. de la Rosa de Saa, M.Á. Gil, Hypothesis testing-based comparative analysis between rating scales for intrinsically imprecise data, *Int. J. Approx. Reasoning* 88 (2017) 128–147.
- [52] M.A. Lubiano, A. Salas, S. de la Rosa de Saa, M., Montenegro, M.Á. Gil, An empirical analysis of the coherence between fuzzy rating scale- and Likert scale-based responses to questionnaires. In: M.B. Ferraro, P. Giordani, B. Vantaggi, M. Gagolewski, M.Á. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), *Soft Methods for Data Science. Advances in Intelligent Systems and Computing*, 456, Springer, Heidelberg, 2017, pp. 329–337.
- [53] M.A. Lubiano, A. Salas, M.Á. Gil, A hypothesis testing-based discussion on the sensitivity of means of fuzzy data with respect to data shape, *Fuzzy Sets Syst.* 328 (2017) 54–69.

- [54] D. McNeish, Thanks coefficient alpha, we'll take it from here, *Psychol. Methods* 23(3) (2018) 412–433.
- [55] J.M. Mendel, M.M. Korjani, A new method for calibrating the fuzzy sets used in fsQCA, *Inf. Sci.* 468 (2018) 155–171.
- [56] N. Menold, C. Wolf, K. Bogner, Design aspects of rating scales in questionnaires, *Math. Popul. Stud.* 25(2) (2018) 63–65.
- [57] R.E. Moore, *Interval Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1966.
- [58] I.A. Motawa, C.J. Anumba, A. El-Hamalawi, A fuzzy system for evaluating the risk of change in construction projects, *Adv. Eng. Softw.* 37 (2006) 583–591.
- [59] E.N. Nasibov, A. Mert, On methods of defuzzification of parametrically represented fuzzy numbers, *Autom. Control Comp. Sci.* 41(5) (2007) 265–273.
- [60] J.C. Nunnally, *Psychometric Methods*, McGraw-Hill, New York, 1978.
- [61] J.C. Nunnally, H. Bernstein, *Psychometric Theory*, McGraw Hill, New York, 1994.
- [62] C.E. Osgood, G. Suci, P. Tannenbaum, *The Measurement of Meaning*, University of Illinois Press, Urbana, 1957.
- [63] M.L. Puri, D.A. Ralescu, Fuzzy random variables, *J. Math. Anal. Appl.* 114 (1986) 409–422.
- [64] P. Quirós, J.M. Alonso, D.P. Pancho, Descriptive and comparative analysis of human perceptions expressed through fuzzy rating scale-based questionnaires, *Int. J. Comput. Intell. Syst.* 9(3) (2016) 450–467.
- [65] T. Raykov, G.A. Marcoulides, Thanks coefficient alpha, we still need you!, *Educ. Psychol. Meas.* 79(1) (2019) 200–210.
- [66] M.A. Robinson, Using multi-item psychometric scales for research and practice in human resource management, *Hum. Resour. Manage.* 57(3) (2018) 739–750.
- [67] E.H. Ruspini, Numerical methods for fuzzy clustering, *Inf. Sci.* 2(3) (1970) 319–350.
- [68] S.K. Savaş, E. Nasibov, A fuzzy ID3 induction for linguistic data sets, *Int. J. Intell. Syst.* 33(4) (2018) 858–878.
- [69] F. Şahin, H. Karadağ, B. Tuncer, Big five personality traits, entrepreneurial self-efficacy and entrepreneurial intention. A configurational approach, *Int. J. Entrep. Behav. Res.* 25(6) (2019) 1188–1211.
- [70] B. Sinova, M.Á. Gil, A. Colubi, S. Van Aelst, The median of a random fuzzy number. The 1-norm distance approach, *Fuzzy Sets Syst.* 200 (2012) 99–115.
- [71] M. Smithson, Fuzzy Sets and Fuzzy Logic in the Human Sciences. In: C. Kahraman, U. Kaymak, A. Yazici (Eds.) *Fuzzy Logic in Its 50th Year, Studies in Fuzziness and Soft Computing*, 341, Springer, Heidelberg, 2016, pp. 175–186.
- [72] S. Srivastava, M. Pant, N. Agarwal, A review on role of Fuzzy Logic in Psychology. In: M. Pant, K. Deep, J.C. Bansal, A. Nagar, K.N. Das (Eds.) *Proceedings of Fifth International Conference on Soft Computing for Problem Solving, Advances in Intelligent Systems and Computing* 437, Springer, Heidelberg, 2016, pp. 783–794.
- [73] L. Stefanini, B. Bede, Generalized fuzzy differentiability with LU-parametric representation, *Fuzzy Sets Syst.* 257 (2014) 184–203.
- [74] L. Stefanini, L. Sorini, M.L. Guerra, Parametric representation of fuzzy numbers and applications to fuzzy calculus, *Fuzzy Sets Syst.* 157 (2006) 2423–2455.
- [75] J. Stoklasa, T. Talášek, J. Musilová, Fuzzy approach - a new chapter in the methodology of Psychology?, *Human Affairs* 24 (2014) 189–203.

- [76] A.L. Sullivan, S. Sadeh, A.K. Hour, Are school psychologists' special education eligibility decisions reliable and unbiased?: A multi-study experimental investigation, *J. Sch. Psychol.* 77 (2019) 90–109.
- [77] Y.-T. Sung, J.-S. Wu, The Visual Analogue Scale for rating, ranking and paired-comparison (VAS-RRP): A new technique for psychological measurement, *Behav. Res. Methods* 50 (2018) 1694–1715.
- [78] K. Takemura, Ambiguous comparative judgment: Fuzzy set model and data analysis, *Jpn. Psychol. Res.* 49(2) (2007) 148–156.
- [79] L.L. Thurstone, Theory of attitude measurement, *Psychol. Rev.* 36(3) (1929) 222–241.
- [80] J. Tijmstra, M. Bolsinova, M. Jeon, General mixture item response models with different item response structures: Exposition with an application to Likert scales, *Behav. Res. Methods* 50(6) (2018) 2325–2344.
- [81] Z.E. Tóth, T. Jónas, R.V. Dénes, Applying flexible fuzzy numbers for evaluating service features in healthcare - patients and employees in the focus, *Total Qual. Manag. Bus. Excell.* 30(S1) (2019) S240-S254.
- [82] W. Trutschnig, M.A. Lubiano, SAFD: Statistical Analysis of Fuzzy Data (2019) Version 2.1.
- [83] J. Verkuilen, T.D. Clark, C.N. Price, A.M. Racanello, Fuzzy Set Theory, SAGE Research Methods Foundations, Foundation Entry, 2020.
- [84] K. Verswijvel, W. Heirman, K. Hardies, M. Walrave, Designing and validating the friendship quality on social network sites questionnaire, *Comput. Hum. Behav.* 86 (2018) 289–298.
- [85] Visual Analog Scales, In: B.B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*, Volume 4, SAGE Publications Inc.: Thousand Oaks, CA, 2018.
- [86] U. Wadgave, M.R. Khairnar, Parametric tests for Likert scale: For and against, *Asian J. Psychiatr.* 24 (2016) 67–68.
- [87] E. Wetzel, S. Greiff, The world beyond rating scales. Why we should think more carefully about the response format in questionnaires, *Eur. J. Psychol. Assess.* 34 (2018) 1–5.
- [88] R.R. Yager, A procedure for ordering fuzzy subsets of the unit interval, *Inf. Sci.* 24 (1981) 143–161.
- [89] C.H. Yeh, R.J. Willis, H. Deng, H. Pan, Task oriented weighting in multi-criteria analysis, *Eur. J. Oper. Res.* 119 (1999) 130–146.
- [90] A.W.K. Yeung, N.S.M. Wong, The historical roots of Visual Analog Scale in Psychology as revealed by reference publication year spectroscopy, *Front. Hum. Neurosci.* 13 (2019) 86.
- [91] H. Yurdugül, Minimum sample size for Cronbach's coefficient alpha: a Monte Carlo study, *Hacettepe Egitim Dergisi* 35 (2008) 397–405.
- [92] L.A. Zadeh, Fuzzy sets, *Inf. Cont.* 8 (1965) 338–353.
- [93] L.A. Zadeh, Quantitative fuzzy semantics, *Inf. Sci.* 3 (1971) 159–176.
- [94] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning. Part 1, *Inf. Sci.* 8 (1975) 199–249; Part 2, *Inf. Sci.* 8 (1975) 301–353; Part 3, *Inf. Sci.* 9 (1975) 43–80.
- [95] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, *Comput. Math. Appl.* 9 (1983) 149–184.
- [96] L.A. Zadeh, Is there a need for fuzzy logic?, *Inf. Sci.* 178 (2008) 2751–2779.

- [97] T. Zétényi (Ed.), Fuzzy Sets in Psychology, Advances in Psychology, 56, North-Holland/Elsevier, Amsterdam, 1988.
- [98] Q. Zhou, Z. Xu, N.Y. Yen, User sentiment analysis based on social network information and its application in consumer reconstruction intention, Comput. Hum. Behav. 100 (2019) 177–183.