# Hypothesis testing-based comparative analysis between rating scales for intrinsically imprecise data ⋆

María Asunción Lubiano*, Antonia Salas, Carlos Carleos, Sara de la Rosa de Sáa, María Ángeles Gil

*Dpto. de Estadística e I.O. y D.M., Universidad de Oviedo, 33007 Oviedo, Spain*

**Abstract**

In previous papers, it has been empirically proved that descriptive (summary measures) and inferential conclusions (in particular, tests about means $p$-values) with imprecise-valued data are often affected by the scale considered to model such data. More concretely, conclusions from the numerical and fuzzy linguistic encodings of Likert-type data have been compared with those for fuzzy data obtained by using a totally free fuzzy assessment: the so-called fuzzy rating scale. These previous comparisons have been performed separately for each of the scales.

This paper aims to perform a joint comparison in such a way that means of linked data (one associated with the fuzzy rating and the other one with the encoded Likert scale) are to be tested for equality. Two real-life examples, as well as several simulation-based synthetic ones, have unequivocally shown that the fuzzy rating scale means are significantly different from those for the encoded Likert scales.

*Key words:* fuzzy linguistic scale; fuzzy rating scale; intrinsically imprecise data; Likert-type scale; testing hypothesis about means

## 1 Introduction

Fuzzy rating scales were introduced (see Hesketh *et al.* [20]) as a computationally/mathematically handleable and expressive tool to rate intrinsically imprecise-valued magnitudes mostly associated with human judgments. By intrinsically imprecise-valued magnitudes we mean in this paper those for which values cannot be in general expressed by means of real numbers, but they can be properly formalized by means of fuzzy numbers.

⋆ This paper is dedicated to the memory of our beloved and admired scientific ancestor, colleague and friend, Professor Pedro Gil. He suggested us a long time ago to analyze the combining of Statistics and Fuzzy Logic. Thanks for your care, Pedro!

* Corresponding author. Tel:+34985103190; fax:+34985103354, lubiano@uniovi.es

The most popular scales to rate such magnitudes are Likert-type ones. They allow a rater to choose among a small number of pre-specified ‘linguistic values’, labeling different degrees of agreement/satisfaction/accomplishment/etc., the one that best represents rater’s score. To develop statistics with Likert scale-based data, the usual way to proceed is to numerically encode different Likert scale values (frequently, consecutive integers), so that the imprecision of Likert values is lost in most cases.

Aiming to capture such an imprecision in a computationally/mathematically (and, hence, statistically) handleable way, a fuzzy linguistic variable, or its associated fuzzy linguistic scale, is introduced (Zadeh [37]) as a fuzzy number-valued encoding of a Likert-type scale.

Fuzzy rating scales confer an added value to fuzzy linguistic ones, namely, the freedom in rating. This freedom in rating results in a much richer and more expressive information, so that diversity, variability and subjectiveness are also much better captured with the fuzzy rating scales than with the Likert or the fuzzy linguistic ones.

Intuitively, because of such a freedom and since fuzzy sets offer more flexibility in an opinion expression, the fuzzy rating scales are more informative than the others from a statistical perspective. In fact, statistical conclusions, should substantially differ depending on the involved scale. This assertion has been recently confirmed from both descriptive and inferential studies (see de la Rosa de Sáa *et al.* [10], Gil *et al.* [15], Lubiano *et al.* [22,23]). In these studies, different descriptive analyses and hypothesis tests about means have been separately developed for each of the three rating scales. Outputs have been later compared leading to conclude that in many of the considered cases they differ to a greater or lesser extent.

In this paper, a joint comparative hypothesis testing-based discussion is carried out. Thus, instead of developing separate tests about means for the three rating scales (as in Lubiano *et al.* [23]), two-sample test about means are to be performed where one of the samples corresponds to fuzzy rating scale-based data and the other one corresponds to either numerically- or fuzzy linguistic-encoded Likert-based data. To avoid the possible influence of raters in the comparative discussion, samples have been taken to be linked.

To get general theoretical conclusions for this comparative discussion would be a chimera. We could always think about a rather unrealistic artificial example leading to conclude that the mean of fuzzy rating scale-based data is not significantly different from that of either numerically- or fuzzy linguistic-encoded Likert-based ones at most of the usual significance levels. Consequently, the approach to be followed is to combine empirical and simulation researches. More concretely, two case studies involving double-type responses (fuzzy rating and Likert) will illustrate the assertion that the mean of fuzzy rating scale-based data is significantly different from that of either numerically- or fuzzy linguistic-encoded Likert-based ones at most of the usual significance levels. This assertion will be more widely corroborated by means of simulations mimicking real-life situations as well as reasonable double-type responses.

In Section 2 of this paper we describe the three scales to be compared. Section 3 recalls the main mathematical tools to be considered for the comparison. Section 4 shows through two real-life examples that in most of the cases the means for the fuzzy rating scale-based random elements significantly differ from those of the associated numerical/fuzzy linguistic-encoding of Likert-based random elements. This conclusion is confirmed in Section 5 by considering simulation developments. The paper ends with some final remarks.

## 2 Preliminaries on the scales to rate intrinsically imprecise-valued magnitudes

This section aims to review the three scales to rate intrinsically imprecise-valued magnitudes we have previously referred to: Likert-type scales (along with their numerical encoding), fuzzy linguistic scales and the fuzzy rating scales.

### *2.1 Likert scale-based ratings*

Likert scale-based ratings [21], allow a rater to choose among a small number of pre-specified 'linguistic values', labeling different degrees of agreement/satisfaction/fulfillment/accomplishment/etc., the one that best represents rater's score.

Figure 1 displays two Likert scale-based items drawn from two real-life questionnaires to be later detailed.

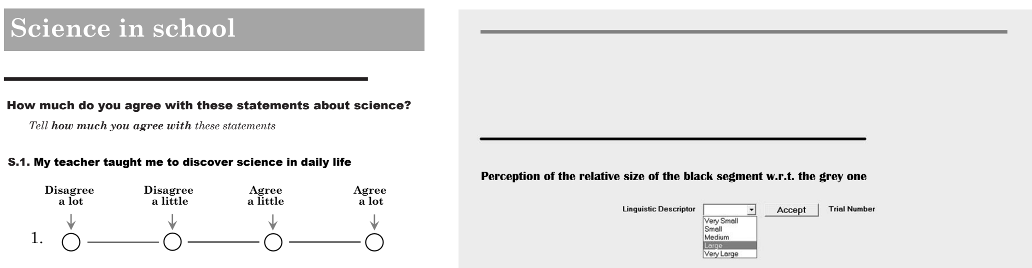



Fig. 1. Examples of 4-point (on the left) and 5-point (on the right) Likert scale-based items from two questionnaires

The item on the left of Figure 1 has been taken from the well-known TIMSS-PIRLS 2011 student questionnaire which is conducted on Grade 4 students (nine to ten years old at the moment they fill the questionnaire) and concerns their opinion and feeling on aspects regarding reading, math, and science. This questionnaire (http://timssandpirls.bc.edu/pirls2011/downloads/P11_StuQ.pdf, and http://timssandpirls.bc.edu/timss2011/downloads/T11_StuQ_4.pdf) is a rather standard paper-and-pencil questionnaire and most of the involved questions have to be answered according to a 4-point Likert scale, responses (linguistic values to choose among) being DISAGREE A LOT, DISAGREE A LITTLE, AGREE A LITTLE, and AGREE A LOT).

The item on the right of Figure 1 has been taken from an online (computerized) application (http://bellman.ciencias.uniovi.es/SMIRE/Perceptions.html) asking users for their perception of the relative length of different line segments (in black) with respect to a longer reference line (in gray). This question has to be answered according to a 5-point Likert scale, responses (linguistic values to choose among) being VERY SMALL, SMALL, MEDIUM, LARGE, and VERY LARGE).

Among the *pros* of using Likert scales one can highlight the following:

- the ease of rating, irrespectively of the framework in which the rating is carried out;
- there is no need for a special training to use them, since common sense is generally enough; as a consequence, Likert scale-based ratings can be usually conducted irrespectively of the age, background, knowledge... of raters;
- the linguistic labels are coherent with the intrinsic imprecision associated with the rating based on these scales.

Among the *cons* that have been pointed out in the literature, one can mention the following:

- the number of possible ‘values’ to choose among is small (i.e., Likert scales are discrete with a small cardinal) and should be usually chosen beforehand; consequently, the variability/adjustment/diversity/subjectivity of these ratings cannot be well captured with these scales;
- the choice of the ‘value’ that best represents rater’s score is often a complex task because none of them accurately fit such a score;
- to analyze Likert-type data a posterior numerical-encoding of the involved Likert scale ‘values’ is usually considered; as a consequence, Likert scale-based data are often treated and analyzed as ordinal (by encoding them by means of their position in accordance with a certain ranking); this makes all differences between consecutive ‘values’ to coincide, which is often seen as inappropriate;
- the transition from a value to another within the scale is rather abrupt;
- the number of suitable statistical techniques to analyze Likert-type data is quite limited, and they are mainly based on the frequencies of different ‘values’ and, maybe, on their numerical encoding, whence relevant statis-

tical information is often lost; actually, many of the commonly employed statistical procedures, albeit applicable, are not really appropriate to deal with Likert-type data.

### *2.2 Fuzzy number scale-based ratings*

The preceding drawbacks lead us to a rather natural question: why not fuzzy scales to rate intrinsically imprecise magnitudes? In the literature one can find several quotations motivating and supporting this endeavour, like "... The fuzzy scales establish a link between strongly defined measurements... and weakly defined measurements" (see Benoit [3]).

Fuzzy scales can be applied to overcome the limitations of standard scales to rate intrinsically imprecise magnitudes by modeling such an imprecision in terms of fuzzy numbers so that

– values capture 'differences in location',
– values capture 'differences in imprecision',
– and they can be mathematically treated.

Fuzzy numbers (also referred to by some authors as fuzzy intervals) are formalized as follows:

**Definition 2.1** *A (bounded)* ***fuzzy number*** *is a function* $\widetilde{U} : \mathbb{R} \to [0,1]$ *such that it is upper semi-continuous, quasi-concave, normal (i.e., it takes on the value 1 for at least a real number), and its support is a bounded interval. In this view (often referred to as the vertical definition), for each* $x \in \mathbb{R}$*, the value* $\widetilde{U}(x)$ *can be interpreted as the 'degree of compatibility of* $x$ *with (the property defined by)* $\widetilde{U}$*'. Equivalently, a (bounded)* ***fuzzy number*** *is a mapping* $\widetilde{U} : \mathbb{R} \to [0,1]$ *such that for any* $\alpha \in [0,1]$ *the* $\alpha$*-level set defined as*

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

*with* 'cl' *denoting the closure of the set, is a nonempty compact interval. This equivalent view is often known as the horizontal definition. The space of (bounded) fuzzy numbers will be denoted by* $\mathscr{F}_c^*(\mathbb{R})$*.*

Real numbers and nonempty compact intervals can be viewed as special fuzzy numbers, since each real number $x$ or each nonempty compact interval $I$ can be identified with the indicator function of the corresponding singleton or interval ($\mathbb{1}_{\{x\}}$ and $\mathbb{1}_I$, respectively).

To illustrate the idea of fuzzy number one can consider a well-known and frequently used family of fuzzy numbers: trapezoidal fuzzy numbers. If $a, b, c, d$

∈ $\mathbb{R}$ with $a \le b \le c \le d$, the trapezoidal fuzzy number $\mathrm{Tra}(a,b,c,d)$ is given, in accordance with the vertical view, by

$$\mathrm{Tra}(a,b,c,d)(x) = \begin{cases} (x-a)/(b-a) & \text{if } x \in [a,b) \\ 1 & \text{if } x \in [b,c] \\ (d-x)/(d-c) & \text{if } x \in (c,d] \\ 0 & \text{otherwise} \end{cases}$$

and, in accordance with the horizontal view, and for each $\alpha \in [0,1]$ by

$$\left(\mathrm{Tra}(a,b,c,d)\right)_\alpha = [a+\alpha(b-a), d+\alpha(c-d)].$$

A wider interesting family of fuzzy numbers, including the one of trapezoidal fuzzy numbers, is that of the *LR-fuzzy numbers* (see Dubois and Prade [12]). Recently, it has been empirically shown (see, for instance, Lubiano *et al.* [24]) that the fuzzy means (and also the real-valued variances) of fuzzy number-valued random elements are not significantly affected by the shape chosen to model fuzzy data (that is, by the choice of functions $L$ and $R$).

### *2.2.1 Fuzzy linguistic scales*

A fuzzy linguistic variable (Zadeh [37]), or its associated FUZZY LINGUISTIC SCALE (FLS), is characterized by a 4-tuple $(\mathbf{X}, \mathscr{T}, \mathcal{S}, \mathbb{R})$, where

- $\mathbf{X}$ is the intrinsically imprecise-valued magnitude to be either measured or observed,
- $\mathscr{T}$ is the set of imprecise 'values' of $\mathbf{X}$ (usually referred to as terms),
- $\mathcal{S}$ is the (fuzzy) semantic rule, i.e., a mapping $\mathcal{S} : \mathscr{T} \to \mathscr{F}_c^*(\mathbb{R})$ where $\mathcal{S}(t)$ is the fuzzy number which has been considered to model the imprecise value $t \in \mathscr{T}$.

Figure 2 displays two triangular fuzzy linguistic scales to model labels/responses in Figure 1. These FLS's are the most usual (balanced) semantic representations of the linguistic hierarchies of $k = 4$ (on the left) and $k = 5$ (on the right) levels (see, for instance, Herrera *et al.* [18], Sanz *et al.* [30]).

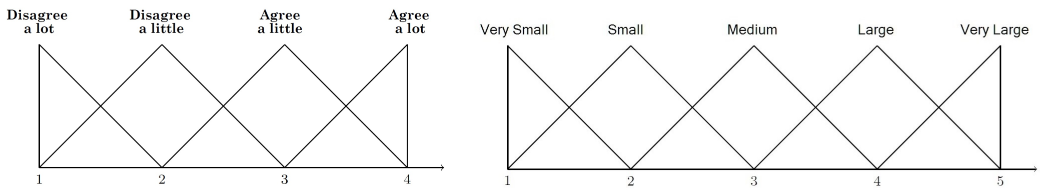



Fig. 2. Examples of 4 terms (on the left) and 5 terms (on the right) fuzzy linguistic scales

**X** corresponds in the first situation to the response chosen for the considered item, and in the second situation it is the perception of the relative length of the shorter line segment w.r.t. the longest reference. Notice that, although in this case there is an underlying precisely-valued magnitude (the real relative length), the perception of such a length (when we are not making use of exact measurement tools) is essentially imprecise.

FLS's can be very often viewed as a posterior fuzzy number-encoding of a Likert scale, so *pros* and *cons* are quite similar to those for the Likert approach.

Among the *pros* of using fuzzy linguistic scales one can highlight the following:

- the ease of the initial rating and no need for a special training, since the posterior encoding is usually made by trained experts;
- the values in the scale can cope (to some extent) with the intrinsic imprecision associated with this rating.

Among the *cons* to be pointed out, one could mention the following:

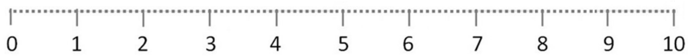
- the number of possible fuzzy values to choose among is small (it is a discrete scale with small cardinal), and the transition from a value to another within the scale is somewhat abrupt; so, variability, adjustment, diversity, subjectivity of these ratings are to some extent lost;
- the choice of the Likert-type 'value' that best represents rater's score is often a complex task because none of them accurately fit such a score, and the same usually happens with the fuzzy modeling of the chosen value; actually, the analyst or another expert transforms the Likert scale labels into fuzzy numbers by choosing a set of fuzzy numbers that he/she finds appropriate to reflect the underlying imprecision in the recorded Likert scale measurements; but this is an arbitrary choice which may or may not reflect the imprecision in the opinion of the persons who originally filled in the questionnaire;
- statistical techniques should be developed to analyze fuzzy number-valued data; in fact, this is currently a rather minor concern, since it has been overcome in the last years, as it will be commented in Section 3.

#### *2.2.2 Fuzzy rating scales*

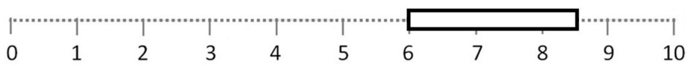
Several quotations from the literature have accurately captured the spirit behind fuzzy rating scales. Among them, we have chosen two that properly motivate and illustrate the aim and scope of these scales: *"... a scale in which... something can be meaningful although we cannot name it"* (Ghneim [14]), and *"Paradoxically, one of the principal contributions of fuzzy logic,... , is its high power of 'precisiation' of what is imprecise"* (see Zadeh [38]).

A FUZZY RATING SCALE (FRS), as introduced by Hesketh *et al.* [20], allows a rater to draw the fuzzy number that 'best represents' rater's score. The guideline for the mechanism to draw such a fuzzy number is as follows:

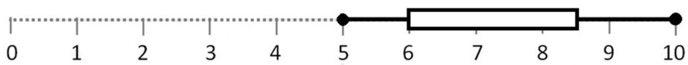
***Step 1.*** A reference bounded interval/segment is first considered. This is often chosen to be $[0, 10]$ or $[0, 100]$, but the choice of such a reference interval is only constrained to be bounded. The end-points are often labeled in accordance with their meaning referring to the degree of agreement, satisfaction, quality, and so on.



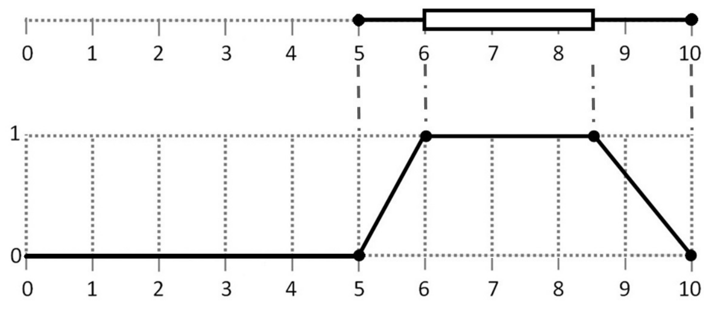

***Step 2.*** The *core*, or 1-level set, associated with the response is determined. It corresponds to the interval consisting of the real values within the reference one which are considered to be as 'fully compatible' with the response.




***Step 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, and it should be always included in the reference interval.




***Step 4.*** The two intervals are 'linearly interpolated' to get a trapezoidal fuzzy number.




It should be pointed out that the linearity of the last interpolation is not a must but it is simply very convenient for computational purposes.

Among the *pros* of using fuzzy rating scales one can highlight the following:

– values in FRS's can cope (to a full extent) with the intrinsic imprecision associated with this rating;

- any FRS means a continuum, and the transition from a value to another within the scale is fully gradual (both in location and precision);
- these scales are much richer and more expressive than any one based on a (unavoidably finite) natural language or its real/fuzzy-valued encoding ("... something can be meaningful although we cannot name it");
- the flexibility of FRS's allow raters to properly capture individual differences, whence the intrinsic variability, diversity and subjectivity are better caught ("... precisiation of what is imprecise");
- values in the FRS's can be mathematically and computationally handled in a suitable way, since one can state arithmetic and distances
  - preserving the meaning of fuzzy numbers,
  - and allowing us to extend/adapt/develop many concepts and developments from Statistics with real-valued data.

Among the *cons* to be pointed out, one could mention the following:

- surveys/questionnaires for which responses are based on a FRS cannot be conducted in any framework, since they require either a paper-and-pencil or a computerized form to be filled by the rater;
- raters need either to have an adequate background or to be properly trained; it should be remarked that, although this is a clear concern, the training does not need to be highly time consuming in most of the cases, as it will be shown in the first case study to be considered in Section 4;
- statistical techniques should be developed to analyze fuzzy number-valued data; in this respect, Hesketh *et al.* [19] have stated 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"; as it has been already commented, nowadays this is just partially a *con*.

From a data-analytic perspective, it is intuitively clear that FRS-based data are much more informative than Likert-based ones (or their numerical/fuzzy encodings). This is due to the fact that, in case data can be doubly rated following both scales, many data matching for the Likert-type scale (and hence showing no variability) do not match at all for the fuzzy rating one (and hence showing a certain variability). To support and illustrate this last assertion, we can consider an item from the case study to be detailed in Section 4.2, in which some items from the TIMSS-PIRLS 2011 student questionnaire have been adapted to allow a double-type response: the original Likert and a FRS-based one with reference interval $[0, 10]$; for instance, in responding to item $M.2$: 'My math teacher is easy to understand' the Likert scale-based response chosen by four students has corresponded to DISAGREE A LITTLE, whereas the FRS-based responses for the same students have been definitely different (see Figure 3).

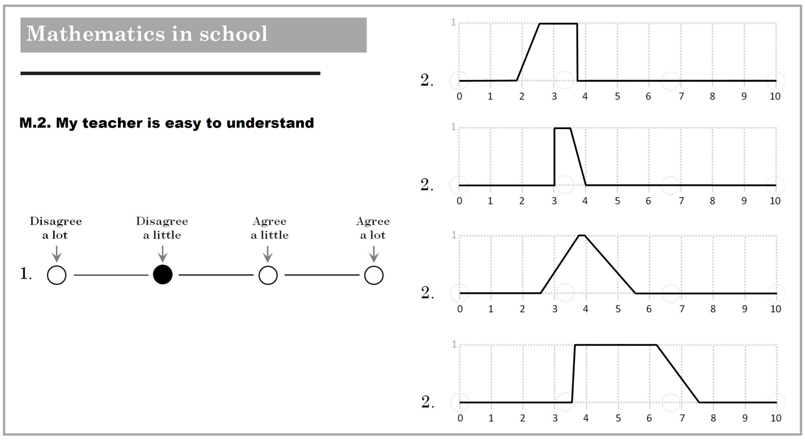



Fig. 3. Example of 4 double responses to item $M.2$ for which the Likert-type ones coincide while the fuzzy rating scale-type clearly differ

From a philosophical perspective, we can wonder about the internal and external consistencies of the FRS-based data. Regarding the internal consistency, if (for instance) essentially the same question is asked repeatedly under similar circumstances to the same person, he/she could often give different answers because of the 'continuous' freedom in drawing such answers; but these answers are expected not to be very different/distant (i.e., they are expected to express a rather consistent opinion), so that almost generally such minor differences scarcely affect the statistical conclusions. Regarding the external consistency, although different subjects sharing the same opinion in connection with a question could express their answers by means of different fuzzy numbers, these differences will be usually much lower than different answers based on a Likert scale; this situation mainly arises when the shared opinion corresponds in fact to an answer that cannot fit any of the Likert scale values but something in between two of them, and in such a situation FRS-based responses will mostly differ less than Likert-type ones, whence statistical conclusions will not be very much affected. Anyway, subjectivity would be unavoidable because of the intrinsic imprecision associated with the aspects to be measures/answered.

It should be noticed that other popular scales which has been used in rating imprecise-valued magnitudes (like pain and many others, coming often from the medical realm) are visual analogue ones, introduced by Freyd [13]. They allow a rater to draw/choose within a given bounded interval (with labeled extremes) the point that best represents rater's score. It shares the *cons* with the FRS, but it also add some specific concerns, namely, that the choice of the single real number that best represents rater's score is usually neither easy nor natural, and to require a full accuracy seems rather unrealistic in such an intrinsically imprecise context.

Regarding the last *con* in connection with the two described fuzzy scales (the need for statistical techniques to analyze fuzzy data), it should be pointed out that along the last years a methodology is being developed to statistically analyze fuzzy scale-based data (see Blanco-Fernández *et al.* [5,6] for a recent review and discussions about). Furthermore, an R package is additionally being stated to support its practical implementation (see Trutschnig and Lubiano [33]), so such a *con* has been substantially overcome.

## 3 Preliminaries on the arithmetic, metrics and hypothesis testing methodology to analyze intrinsically imprecise-valued data

The key tools for the above-mentioned statistical methodology with fuzzy data are:

- arithmetic + metrics with fuzzy numbers;
- random fuzzy numbers.

Why does combining arithmetic + metrics constitute a key tool in this setting? To handle fuzzy data from a mathematical perspective, one can first pose a relevant question: can fuzzy data be treated as special functional data? There is not a single answer to the last question, but the two following answers are compatible:

- Directly, NO. In applying functional arithmetic to handle elements in the space of (functional-valued) fuzzy numbers, one often moves out of the space and the fuzzy meaning is generally lost.
- Indirectly, YES. By using an appropriate arithmetic and suitable metrics, fuzzy numbers can be identified with elements in a convex cone of a Hilbert space of functions, and the arithmetic and metrics with fuzzy numbers with those in the Hilbert space of functions (see, for instance, González-Rodríguez *et al.* [16]). This is the view we will adopt along this paper.

### *3.1 Arithmetic with fuzzy data*

When fuzzy numbers are considered to model experimental data, statistics to analyze them are frequently based on two arithmetical operations, namely the sum and the product by scalars.

The common way to extend the sum and the product by a scalar from $\mathbb{R}$ to $\mathscr{F}_c^*(\mathbb{R})$ is to use Zadeh's extension principle [37], which is equivalent to considering the usual interval arithmetic level-wise. More concretely,

**Definition 3.1** *Given $\widetilde{U}, \widetilde{V} \in \mathcal{F}_c^*(\mathbb{R})$, the* ***sum*** *of $\widetilde{U}$ and $\widetilde{V}$ is the fuzzy number $\widetilde{U} + \widetilde{V} \in \mathscr{F}_c^*(\mathbb{R})$ such that for each $\alpha \in [0, 1]$*

$$(\widetilde{U} + \widetilde{V})_\alpha = \widetilde{U}_\alpha + \widetilde{V}_\alpha = \left[\inf \widetilde{U}_\alpha + \inf \widetilde{V}_\alpha, \sup \widetilde{U}_\alpha + \sup \widetilde{V}_\alpha\right].$$

*Given $\widetilde{U} \in \mathscr{F}^*_c(\mathbb{R})$ and $\gamma \in \mathbb{R}$, the* ***product*** *of $\widetilde{U}$* ***by the scalar*** *$\gamma$ is the fuzzy number $\gamma \cdot \widetilde{U} \in \mathscr{F}^*_c(\mathbb{R})$ such that for each $\alpha \in [0,1]$*

$$(\gamma \cdot \widetilde{U})_\alpha = \gamma \cdot \widetilde{U}_\alpha = \begin{cases} \left[\gamma \cdot \inf \widetilde{U}_\alpha, \gamma \cdot \sup \widetilde{U}_\alpha\right] & \text{if } \gamma \geq 0 \\ \left[\gamma \cdot \sup \widetilde{U}_\alpha, \gamma \cdot \inf \widetilde{U}_\alpha\right] & \text{otherwise.} \end{cases}$$

**Remark 3.1** It can be easily proved that both operations are closed within the class of trapezoidal fuzzy numbers.

**Remark 3.2** It should be especially highlighted that the space $(\mathscr{F}^*_c(\mathbb{R}), +, \cdot)$ has not linear but semilinear structure since $\widetilde{U} + (-1 \cdot \widetilde{U}) \neq \mathbb{1}_{\{0\}}$ (neutral element of $+$).

### *3.2 Metric between fuzzy data*

Due to the nonlinearity that has been pointed out in Remark 3.2, one cannot state a definition for the difference between fuzzy numbers that is always well-defined and simultaneously preserves the main properties of the difference between real values in connection with the sum. In fact, there exists a difference notion (Hukuhara's one) satisfying the last condition, but it cannot be defined for many fuzzy number values.

This crucial drawback has been substantially overcome in developing statistics with fuzzy data by incorporating suitable distances between them. On one hand, distances will allow to 'translate' the equality of fuzzy numbers into the vanishing of the distance between them, as in the case of real values. On the other hand, appropriate distances also allow us *via* the support function to 'identify' fuzzy data with functional ones and fuzzy arithmetic with functional arithmetic (as it will be later remarked). Furthermore, statistical concepts and methods for real-valued datasets involving metrics (e.g., dispersion measures, mean distance approaches, classification problems, etc.) could be extended by considering extended metrics.

Among the $L^2$ metrics between fuzzy numbers, the one introduced by Diamond and Kloeden [11], and extending Vitale's [34] one for interval values, is given as follows:

**Definition 3.2** *Let $\widetilde{U}, \widetilde{V} \in \mathscr{F}^*_c(\mathbb{R})$. The* ***2-norm distance*** *between $\widetilde{U}$ and $\widetilde{V}$ is defined as*

$$\rho_2(\widetilde{U}, \widetilde{V}) = \sqrt{\frac{1}{2}\int_{(0,1]} \left(\left[\inf \widetilde{U}_\alpha - \inf \widetilde{V}_\alpha\right]^2 + \left[\sup \widetilde{U}_\alpha - \sup \widetilde{V}_\alpha\right]^2\right) d\alpha}$$

$$= \sqrt{\int_{[0,1]} \left([\operatorname{mid} \widetilde{U}_\alpha - \operatorname{mid} \widetilde{V}_\alpha]^2 + [\operatorname{spr} \widetilde{U}_\alpha - \operatorname{spr} \widetilde{V}_\alpha]^2\right) d\alpha},$$

*where* mid *and* spr *are the centre and radius, respectively, of the corresponding interval (i.e.,* $\operatorname{mid}\widetilde{U}_\alpha = (\inf \widetilde{U}_\alpha + \sup \widetilde{U}_\alpha)/2$*,* $\operatorname{spr}\widetilde{U}_\alpha = (\sup \widetilde{U}_\alpha - \inf \widetilde{U}_\alpha)/2$*).*

**Remark 3.3** In dealing with trapezoidal fuzzy numbers and $\rho_2$, we have that

$$\rho_2(\mathrm{Tra}(a_1, b_1, c_1, d_1), \mathrm{Tra}(a_2, b_2, c_2, d_2))$$

$$= \sqrt{\frac{(a_1-a_2)^2 + (b_1-b_2)^2 + (a_1-a_2)(b_1-b_2) + (c_1-c_2)^2 + (d_1-d_2)^2 + (c_1-c_2)(d_1-d_2)}{6}}$$

$$= \sqrt{\frac{(\underline{m}_1-\underline{m}_2)^2 + (\overline{m}_1-\overline{m}_2)^2 + (\underline{m}_1-\underline{m}_2)(\overline{m}_1-\overline{m}_2) + (\underline{r}_1-\underline{r}_2)^2 + (\overline{r}_1-\overline{r}_2)^2 + (\underline{r}_1-\underline{r}_2)(\overline{r}_1-\overline{r}_2)}{3}},$$

where $\underline{m} = (a+d)/2$, $\overline{m} = (b+c)/2$, $\underline{r} = (d-a)/2$ and $\overline{r} = (b-c)/2$.

**Remark 3.4** By combining the above fuzzy arithmetic and metric $\rho_2$, and *via* the so-called support function introduced by Puri and Ralescu [28], $s : \mathscr{F}_c^*(\mathbb{R}) \to \mathbb{H}_2$ (with $\mathbb{H}_2 = \{L^2$-type real-valued functions defined on $[0,1] \times \{-1,1\}$ w.r.t. $\ell \otimes \lambda_1\}$, $\lambda_1(-1) = \lambda_1(1) = .5$, and $s(\widetilde{U}) = s_{\widetilde{U}}$ with $s_{\widetilde{U}}(\alpha, -1) = -\inf \widetilde{U}_\alpha$, $s_{\widetilde{U}}(\alpha, 1) = \sup \widetilde{U}_\alpha$), an isometric embedding of $\mathscr{F}_c^*(\mathbb{R})$ onto a convex cone of the Hilbert space $\mathbb{H}_2$ can be stated. An immediate and crucial implication from such an embedding is that any fuzzy number $\widetilde{U} \in \mathscr{F}_c^*(\mathbb{R})$ can be identified with the corresponding function $s_{\widetilde{U}}$ and this identification is accompanied by the correspondences between the usual arithmetics and $L^2$ metrics. Consequently, data in the setting of fuzzy number-valued data with the fuzzy arithmetic and the metric $\rho_2$ (in fact, with more general $L^2$ metrics, see González-Rodríguez *et al.* [16]) can be systematically translated into data in the setting of functional data with the functional arithmetic and the metric based on the associated norm. In this way, fuzzy data should not be treated directly, but *via* the support function, as functional data.

Then, we can formally assert as a relevant implication for statistical purposes that several developments in Functional Data Analysis could be particularized to fuzzy number-valued data by using the adequate identifications and correspondences. However, it should be guaranteed that the resulting elements/outputs and steps remain in the cone $s\,(\mathscr{F}_c^*(\mathbb{R}))$. This will apply, for instance, in testing about means in Section 3.4.

### *3.3 Random fuzzy numbers*

In developing statistics with fuzzy data coming from intrinsically imprecise-valued magnitudes, random fuzzy numbers constitute a well-formalized model within the probabilistic setting for the random mechanisms generating such data. Random fuzzy numbers, originally coined as (one-dimensional) fuzzy random variables by Puri and Ralescu [29], integrate randomness (associated with data generation) and fuzziness (associated with data nature).

**Definition 3.3** *Let $(\Omega, \mathcal{A}, P)$ be a probability space modeling a random experiment. A mapping $\mathcal{X} : \Omega \to \mathscr{F}_c^*(\mathbb{R})$ is said to be an associated* ***random fuzzy number*** *(for short RFN) if and only if for all $\alpha \in [0,1]$ the interval-valued mapping $\mathcal{X}_\alpha$, such that $\mathcal{X}_\alpha(\omega) = \big(\mathcal{X}(\omega)\big)_\alpha$ for all $\omega \in \Omega$, is a compact random interval (i.e., a Borel-measurable mapping w.r.t. the topology induced by Hausdorff metric in the space of the nonempty compact intervals).*

*Equivalently, $\mathcal{X}$ is an RFN if and only if $s(\mathcal{X})$ is an $\mathbb{H}_2$-valued random element (that is, a Borel-measurable function w.r.t. the Borel $\sigma$-field generated by the topology induced by the metric associated with $\rho_2$* via *$s$).*

*Also equivalently, $\mathcal{X}$ is an RFN if and only if it is a Borel-measurable mapping w.r.t. the Borel $\sigma$-field generated on $\mathscr{F}_c^*(\mathbb{R})$ by the topology induced by $\rho_2$.*

**Remark 3.5** The Borel-measurability in the third definition above ensures that one can properly and trivially refer to the *distribution induced by an RFN*, the *stochastic independence of RFN's*, and so on, without needing to state expressly these notions.

In analyzing the induced distribution of a random fuzzy number the best known summary measure is the Aumann-type mean (Puri and Ralescu [29]), that extends the mean of a random variable as well as the Aumann expectation of a random set, and it is formalized as follows:

**Definition 3.4** *Let $\mathcal{X}$ be a random fuzzy number associated with the probability space $(\Omega, \mathcal{A}, P)$. The* ***(population) Aumann-type mean*** *of $\mathcal{X}$ is the fuzzy number $\widetilde{E}(\mathcal{X}) \in \mathscr{F}_c^*(\mathbb{R})$, if it exists, such that for each $\alpha \in [0,1]$*

$$\big(\widetilde{E}(\mathcal{X})\big)_\alpha = \text{Aumann integral of } \mathcal{X}_\alpha = \left\{ \int_\Omega f(\omega)\, dP(\omega) \,:\, f \overset{a.s.\,[P]}{\in} \mathcal{X}_\alpha \right\}$$

*(see Aumann [1]), that is, $\big(\widetilde{E}(\mathcal{X})\big)_\alpha = [E(\inf \mathcal{X}_\alpha), E(\sup \mathcal{X}_\alpha)]$ with $E$ denoting the expected value of a real-valued random variable. Equivalently, and whenever $s_{\mathcal{X}} \in L^1(\Omega, \mathcal{A}, P)$, it is the fuzzy number $\widetilde{E}(\mathcal{X}) \in \mathscr{F}_c^*(\mathbb{R})$ such that $s_{\widetilde{E}(\mathcal{X})} = \mathrm{E}(s_{\mathcal{X}})$, with* E *denoting the Bochner expectation of a Banach space-valued random element.*

**Remark 3.6** In particular, if $\widetilde{\boldsymbol{x}}_{\boldsymbol{n}} = (\mathcal{X}(\omega_1), \ldots, \mathcal{X}(\omega_n))$ is a sample of observations from $\mathcal{X}$ when measured on a sample of individuals $(\omega_1, \ldots, \omega_n)$, the ***(sample) Aumann-type mean*** is the fuzzy number $\overline{\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}}$ given for all $\alpha \in [0,1]$ by

$$\left(\overline{\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}}\right)_\alpha = \left(\frac{1}{n} \cdot (\mathcal{X}(\omega_1) + \ldots + \mathcal{X}(\omega_n))\right)_\alpha = \left[\frac{1}{n}\sum_{i=1}^n \inf(\mathcal{X}(\omega_i))_\alpha, \frac{1}{n}\sum_{i=1}^n \sup(\mathcal{X}(\omega_i))_\alpha\right].$$

**Remark 3.7** If $\mathcal{X}$ is a trapezoidal-valued random fuzzy number then

$$\widetilde{E}(\mathcal{X}) = \mathrm{Tra}\Big(E(\inf \mathcal{X}_0), E(\inf \mathcal{X}_1), E(\sup \mathcal{X}_1), E(\sup \mathcal{X}_0)\Big).$$

In particular, if $\widetilde{\boldsymbol{x}}_{\boldsymbol{n}} = (\mathrm{Tra}(a_1, b_1, c_1, d_1), \dots, \mathrm{Tra}(a_n, b_n, c_n, d_n))$ is a sample of observations from $\mathcal{X}$, the ***(sample) Aumann-type mean*** is the fuzzy number

$$\overline{\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}} = \mathrm{Tra}\left(\frac{1}{n}\sum_{i=1}^{n} a_i, \frac{1}{n}\sum_{i=1}^{n} b_i, \frac{1}{n}\sum_{i=1}^{n} c_i, \frac{1}{n}\sum_{i=1}^{n} d_i\right).$$

The Aumann-type mean preserves the main valuable properties from the real-valued case, so that it is equivariant under affine transformations on $\mathscr{F}_c^*(\mathbb{R})$ (i.e., $\widetilde{E}(a \cdot \mathcal{X} + b) = a \cdot \widetilde{E}(\mathcal{X}) + b$), additive (i.e., $\widetilde{E}(\mathcal{X} + \mathcal{Y}) = \widetilde{E}(\mathcal{X}) + \widetilde{E}(\mathcal{Y})$), coherent with the above-described fuzzy arithmetic (as shown in Remark 3.6), it fulfills Strong Laws of Large Numbers, and it is the Fréchet expectation w.r.t. $\rho_2$ (i.e., $\widetilde{E}(\mathcal{X}) = \arg\min_{\widetilde{U} \in \mathscr{F}_c^*(\mathbb{R})} E\Big[\Big(\rho_2(\mathcal{X}, \widetilde{U})\Big)^2\Big]$).

### *3.4 Two-sample test about means for linked samples of RFN's*

As we have already announced, this paper aims to test that the use of different scales to rate intrinsically imprecise-valued magnitudes can often lead to different statistical conclusions. To confirm this fact, we are going to consider real-life and synthetic examples for which a double simultaneous rating is assumed: a Likert scale- and an FRS-based rating.

Once either real-life or simulated double data are collected or generated, two-sample test about means for linked samples of RFN's are to be performed. More concretely,

- for the considered case studies, Likert-type data are to be encoded both numerically (leading to the so-denoted NELikert-based data) and fuzzy linguistically (leading to the so-denoted FLS-based data); NELikert (actually, the indicator functions of their associated singletons) and FLS data will be treated as fuzzy number-valued data;
- for the considered simulations, NELikert and FLS data will be obtained as the real numbers (again the indicator functions of their associated singletons) or FLS values showing the lowest $\rho_2$-distance to the generated FRS data;
- when all double data are transformed into couples of fuzzy linked data, the null hypothesis about the equality of the corresponding two Aumann-type means is to be tested; in fact, the $p$-value of the two-sample test for linked samples is to be computed.

To test the null hypothesis of equality of the Aumann-type means of two RFNs $\mathcal{X}$ and $\mathcal{X}'$, one can consider the *bootstrapped algorithm* for trapezoidal-valued random fuzzy numbers in Lubiano *et al.* [23] (approximating the particularization of the two-sample test about means for linked samples from RFNs by González-Rodríguez *et al.* [17]).

If $(\mathcal{X}, \mathcal{X}')$ is a two-dimensional random fuzzy set (that is, a mapping from $\Omega$ to $\mathscr{F}_c^*(\mathbb{R}) \times \mathscr{F}_c^*(\mathbb{R})$ for which $\alpha$-levels are compact convex random sets of $\mathbb{R}^2$), consider a sample of independent observations from it, $\big((\widetilde{x}_1, \widetilde{x}'_1), \ldots, (\widetilde{x}_n, \widetilde{x}'_n)\big)$. Assume that $\widetilde{x}_i$ and $\widetilde{x}'_i$ are trapezoidal fuzzy numbers, and denote $\widetilde{\boldsymbol{x}}_{\boldsymbol{n}} = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$, $\widetilde{\boldsymbol{x}}'_{\boldsymbol{n}} = (\widetilde{x}'_1, \ldots, \widetilde{x}'_n)$ with $\widetilde{x}_i = \mathrm{Tra}(a_i, b_i, c_i, d_i)$ and $\widetilde{x}'_i = \mathrm{Tra}(a'_i, b'_i, c'_i, d'_i)$.

Then, the algorithm to test the null hypothesis $H_0 : \widetilde{E}(\mathcal{X}) = \widetilde{E}(\mathcal{X}')$ (i.e., $H_0 : \rho_2\big(\widetilde{E}(\mathcal{X}), \widetilde{E}(\mathcal{X}')\big) = 0$) proceeds as follows:

***Step 1.*** Compute the value of the statistic

$$\mathsf{T}_n(\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}, \widetilde{\boldsymbol{x}}'_{\boldsymbol{n}}) = \frac{\left[\rho_2(\overline{\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}}, \overline{\widetilde{\boldsymbol{x}}'_{\boldsymbol{n}}})\right]^2}{\frac{1}{n}\sum_{i=1}^{n}\left[\rho_2(\widetilde{x}_i + \overline{\widetilde{\boldsymbol{x}}'_{\boldsymbol{n}}}, \widetilde{x}'_i + \overline{\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}})\right]^2} = \frac{A_n(\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}, \widetilde{\boldsymbol{x}}'_{\boldsymbol{n}})}{C_n(\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}, \widetilde{\boldsymbol{x}}'_{\boldsymbol{n}})},$$

$$\begin{aligned} A_n(\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}, \widetilde{\boldsymbol{x}}'_{\boldsymbol{n}}) = &\left[\frac{1}{n}\sum_{i=1}^{n}(\underline{m}_i - \underline{m}'_i)\right]^2 + \left[\frac{1}{n}\sum_{i=1}^{n}(\overline{m}_i - \overline{m}'_i)\right]^2 + \left[\frac{1}{n}\sum_{i=1}^{n}(\underline{r}_i - \underline{r}'_i)\right]^2 + \left[\frac{1}{n}\sum_{i=1}^{n}(\overline{r}_i - \overline{r}'_i)\right]^2 \\ &+ \left[\frac{1}{n}\sum_{i=1}^{n}(\underline{m}_i - \underline{m}'_i)\right] \cdot \left[\frac{1}{n}\sum_{i=1}^{n}(\overline{m}_i - \overline{m}'_i)\right] + \left[\frac{1}{n}\sum_{i=1}^{n}(\underline{r}_i - \underline{r}'_i)\right] \cdot \left[\frac{1}{n}\sum_{i=1}^{n}(\overline{r}_i - \overline{r}'_i)\right]. \end{aligned}$$

$$\begin{aligned} C_n(\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}, \widetilde{\boldsymbol{x}}'_{\boldsymbol{n}}) = &\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}(\underline{m}_i + \underline{m}_l - \underline{m}'_i - \underline{m}'_l)\right]^2 + \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}(\overline{m}_i + \overline{m}_l - \overline{m}'_i - \overline{m}'_l)\right]^2 \\ &+ \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}(\underline{r}_i + \underline{r}_l - \underline{r}'_i - \underline{r}'_l)\right]^2 + \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}(\overline{r}_i + \overline{r}_l - \overline{r}'_i - \overline{r}'_l)\right]^2 \\ &+ \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}(\underline{m}_i + \underline{m}_l - \underline{m}'_i - \underline{m}'_l)\right] \cdot \left[\frac{1}{n}\sum_{l=1}^{n}(\overline{m}_i + \overline{m}_l - \overline{m}'_i - \overline{m}'_l)\right] \\ &+ \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}(\underline{r}_i + \underline{r}_l - \underline{r}'_i - \underline{r}'_l)\right] \cdot \left[\frac{1}{n}\sum_{l=1}^{n}(\overline{r}_i + \overline{r}_l - \overline{r}'_i - \overline{r}'_l)\right]. \end{aligned}$$

***Step 2.*** Fix the bootstrap populations to be $\Big\{\big(\widetilde{x}_1 + \overline{\widetilde{\boldsymbol{x}}'_{\boldsymbol{n}}}, \widetilde{x}'_1 + \overline{\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}}\big), \ldots, \big(\widetilde{x}_n + \overline{\widetilde{\boldsymbol{x}}'_{\boldsymbol{n}}}, \widetilde{x}'_n + \overline{\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}}\big)\Big\}$, with

$$\widetilde{x}_i + \overline{\widetilde{\boldsymbol{x}}'_{\boldsymbol{n}}} = \mathrm{Tra}\left(a_i + \frac{a'_1 + \ldots + a'_n}{n}, b_i + \frac{b'_1 + \ldots + b'_n}{n}, c_i + \frac{c'_1 + \ldots + c'_n}{n}, d_i + \frac{d'_1 + \ldots + d'_n}{n}\right),$$

$$\widetilde{x}'_i + \overline{\widetilde{\boldsymbol{x}}_{\boldsymbol{n}}} = \mathrm{Tra}\left(a'_i + \frac{a_1 + \ldots + a_n}{n}, b'_i + \frac{b_1 + \ldots + b_n}{n}, c'_i + \frac{c_1 + \ldots + c_n}{n}, d'_i + \frac{d_1 + \ldots + d_n}{n}\right)$$

so that to ensure that bootstrap populations fulfill the null hypothesis, one can add to each value in each sample the mean of the other one.

***Step 3.*** Obtain a sample of independent observations from each bootstrap population, say $\{(\widetilde{x}_1, \widetilde{x}'_1)^*, \ldots, (\widetilde{x}_n, \widetilde{x}'_n)^*\}$ and, for the sake of simplicity, denote $(\widetilde{x}^*_i, \widetilde{x}'^*_i) = (\widetilde{x}_i, \widetilde{x}'_i)^*$ and $\widetilde{\boldsymbol{x}}^*_{\boldsymbol{n}} = (\widetilde{x}^*_1, \ldots, \widetilde{x}^*_n)$, $\widetilde{\boldsymbol{x}}'^*_{\boldsymbol{n}} = (\widetilde{x}'^*_1, \ldots, \widetilde{x}'^*_n)$.

***Step 4.*** Compute the value of the bootstrap statistic $\mathsf{T}^*_n(\widetilde{\boldsymbol{x}}^*_{\boldsymbol{n}}, \widetilde{\boldsymbol{x}}'^*_{\boldsymbol{n}})$.

***Step 5.*** *Steps 3* and *4* should be repeated a large number $B$ of times to get a set of $B$ estimates, denoted by $\{\mathsf{t}^*_1, \ldots, \mathsf{t}^*_B\}$.

***Step 6.*** Compute the bootstrap $p$-value as the proportion of values in $\{\mathsf{t}^*_1, \ldots, \mathsf{t}^*_B\}$ being greater than $\mathsf{T}_n$.

## 4 Case studies-based discussion

As it has been already commented, general theoretical conclusions for the equality of means for FRS-based *vs* either NELikert- or FLS-based data cannot be achieved. This section aims to show that means are mostly significantly different in real-life situations. For this purpose, two case studies, one involving a questionnaire with several items and a 4-point Likert-type scale, and the other one involving a single question with several trials and a 5-point Likert type scale, both allowing double-type responses, are to be considered. In the first case, respondents are nine to ten year-old children whereas in the second one respondents are scientists.

### *4.1 Case study 1: adapted TIMSS/PIRLS questionnaire*

This example has been previously examined for different statistical purposes (see Gil *et al.* [15], Lubiano *et al.* [22], Sinova *et al.* [32]). It relates to the well-known questionnaire TIMSS-PIRLS 2011 that has been referred to in explaining Figure 1 (Section 2.1). It has been conducted on the population of Grade 4 students and most of the involved questions have to be answered according to the already described 4-point Likert scale.

To get more expressive responses and informative conclusions, nine items from the original questionnaire have 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 4 for one of the items, http://bellman.ciencias.uniovi.es/SMIRE/FuzzyRatingScaleQuestionnaire-SanIgnacio.html and the supplementary material for the full paper-and-pencil form and http://carleos.epv.uniovi.es:8080/ for the full -Spanish- computerized form).

The nine adapted items chosen from the original Student questionnaire are displayed in Table 1. The adapted questionnaire involving these double-response items has been conducted in 2014 on a sample of 69 fourth grade students from Colegio San Ignacio (Oviedo-Asturias, Spain). These students have been distributed in accordance with (their usual) three groups, and the teachers have

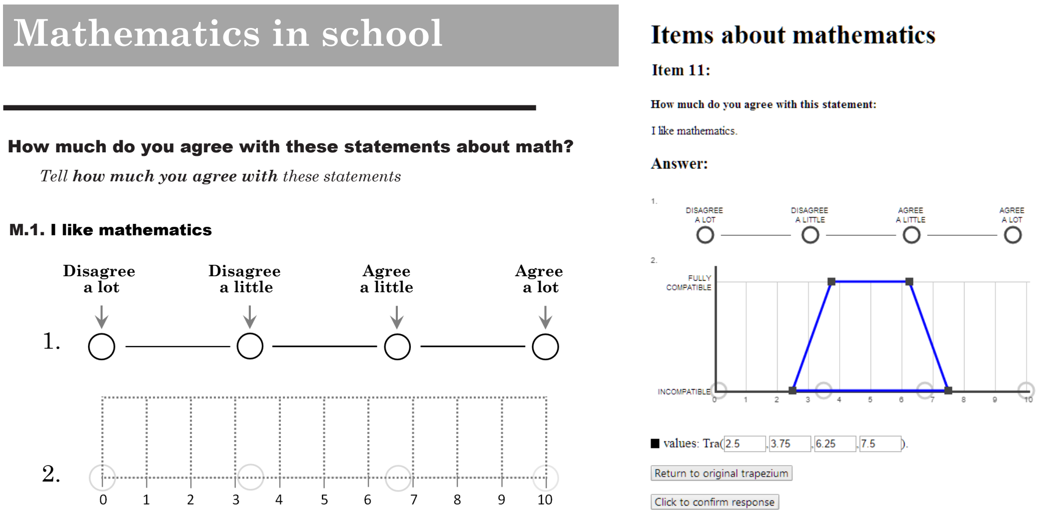



Fig. 4. Example of the double-response paper-and-pencil (on the left) and computerized (on the right) form to an item in Case Study 1

decided that the 24 students in one of the three classrooms have to fill out the paper-and-pencil format and the 45 students from the other two groups have to complete the computerized version. To ‘ease’ the relationship between the two scales for these very young respondents, each numerically encoded Likert response has been lightly superimposed upon the reference interval of the fuzzy rating scale part.

Table 1
Items adapted from the TIMSS-PIRLS 2011 Student’s Questionnaire

| | READING IN SCHOOL |
|---|---|
| $R.1$ | I like to read things that make me think |
| $R.2$ | I learn a lot from reading |
| $R.3$ | Reading is harder for me than any other subject |
| | MATHEMATICS IN SCHOOL |
| $M.1$ | I like mathematics |
| $M.2$ | My teacher is easy to understand |
| $M.3$ | Mathematics is harder for me than any other subject |
| | SCIENCE IN SCHOOL |
| $S.1$ | My teacher taught me to discover science in daily life |
| $S.2$ | I read about in my spare time |
| $S.3$ | Science is harder for me than any other subject |

The training of students to let them know about the meaning and purpose of the case study, as well as the aim of the double response, has been carried out in up to 15 minutes, and three researchers from the Department of Statistics, OR and Math Teaching of the University of Oviedo have been in charge of the explanation and conduction of the survey. At this point, it should be remarked that students had no idea on the concept of real-valued functions and they had just learned that of a trapezium. With the guideline detailed in the supplementary material for this paper, students have not had understanding problems, they have catched the philosophy behind and have been able to provide us with quite coherent responses in most of the cases. Actually, for all the questions, the number of 'no response's' have been very small and smaller for the fuzzy rating than for the Likert scale. In summary, the training has been surprisingly much easier and more effective than it could be expected. Datasets associated with responses to this questionnaire can be also found in the supplementary material.

The bootstrapped two-sample test about means for linked samples in Section 3.4 has been now applied (with $B = 1000$) for each of the nine items in Table 1, with $\mathcal{X}$ standing for the FRS-based response and $\mathcal{X}'$ standing for either the numerically encoded 4-point Likert-based responses (denoted NELikert and taking on values 0, 10/3, 20/3, 10) or the fuzzy linguistically encoded 4-point Likert-based responses in accordance with some of the most frequently FLSs considered when 4 labels are modelled in connection with decision making, classification, control, and other problems for which these scales have shown to be valuable (see, for instance, Herrera *et al.* [18], Bajpai *et al.* [2], Cai *et al.* [7], Picon *et al.* [27]). FLS1 will denote the most usual (balanced) semantic representations of the linguistic hierarchies on the left in Figure 2 (Section 2.2.1). FLS1 to FLS5 are displayed in Figure 5.

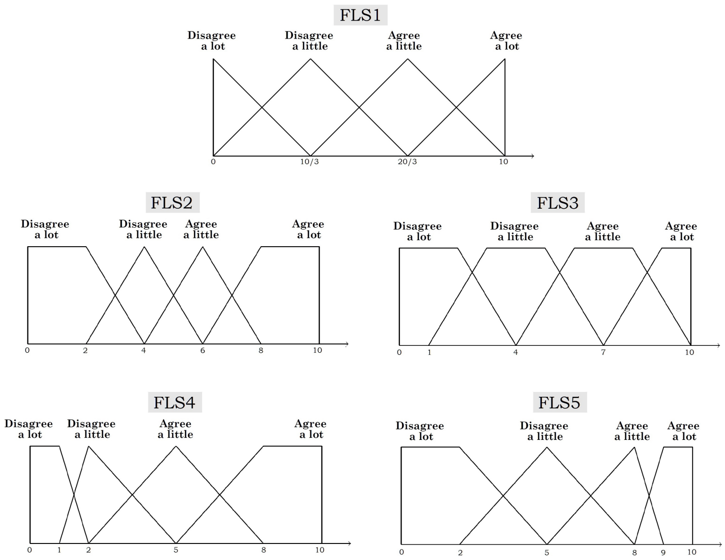



Fig. 5. Examples of five usual fuzzy linguistic scales with 4 terms

Table 2
Bootstrapped $p$-values of the two-sample test about means for linked samples (FRS *vs* encoded scale in {NELikert, FLS1, FLS2, FLS3, FLS4, FLS5})

| item \ $\mathcal{X}'$ | NELikert | FLS1 | FLS2 | FLS3 | FLS4 | FLS5 |
|---|---|---|---|---|---|---|
| $R.1$ | .000 | .000 | .000 | .000 | .000 | .000 |
| $R.2$ | .000 | .000 | .000 | .000 | .000 | .000 |
| $R.3$ | .000 | .016 | .060 | .000 | .000 | .002 |
| $M.1$ | .000 | .000 | .000 | .000 | .000 | .000 |
| $M.2$ | .000 | .002 | .000 | .000 | .000 | .005 |
| $M.3$ | .002 | .000 | .001 | .000 | .006 | .000 |
| $S.1$ | .000 | .000 | .020 | .000 | .000 | .000 |
| $S.2$ | .000 | .000 | .001 | .000 | .000 | .000 |
| $S.3$ | .000 | .000 | .001 | .000 | .000 | .000 |

On the basis of the $p$-values in Table 2 one can almost generally conclude that for any of the nine items and for the most usual significance levels, the FRS-based mean response is significantly different from the encoded Likert-based mean response (when the encoded Likert scale is in {NELikert, FLS1, FLS2, FLS3, FLS4, FLS5}). In summary, the mean response is influenced by the considered scale.

### *4.2 Case study 2: perception of the relative length of a line segment*

This example has been previously examined for different statistical purposes (see Colubi *et al.* [8], González-Rodríguez *et al.* [16]). It relates to an online computerized application in which people have been asked for their perception of the relative length of different line segments with respect to a pattern longer one, and it has been referred to in explaining Figure 1. The population have corresponded to people who can be potentially contacted for this purpose.

The application has been conducted so that on the center top of the screen the longest (reference) line segment has been drawn in gray. This segment is fixed for all the trials, so that there is always a reference for the maximum length. At each trial a black 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.

Each of the perceptions could be doubly expressed, namely by choosing the Likert-like scale in Figure 1, and by using the fuzzy rating scale with reference interval $[0, 100]$ so that they can be thought as a kind of imprecise percentages (see Figure 6 for a screen capture).

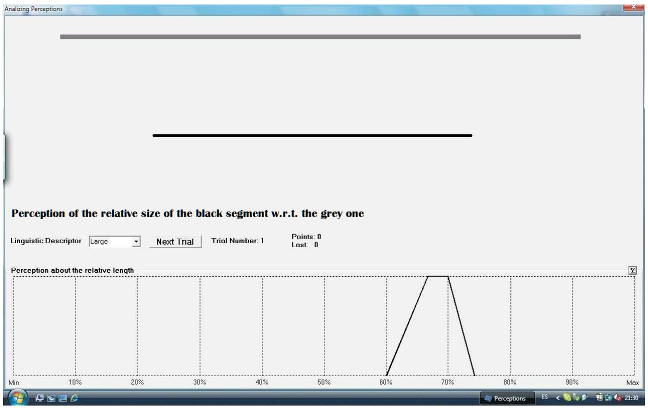



Fig. 6. Example of a double-response computerized question in Case Study 2

The online application explains the formalization and meaning of the fuzzy rating values (see http://bellman.ciencias.uniovi.es/SMIRE/perceptions.html and the supplementary material for this paper).

A sample of 25 respondents (all of them with a university scientific background and with a quite minor training need, mostly consisting of simply reading the instructions in the online application and the supplementary material) have been contacted for this experiment, and they have supplied the responses in the supplementary material for this paper.

The bootstrapped two-sample test about means for linked samples in Section 3.4 has been now applied (with $B = 1000$), with $\mathcal{X}$ standing for the FRS-based response and $\mathcal{X}'$ standing for either the numerically encoded 5-point Likert-based responses (denoted NELikert' and taking on values 0, 25, 50, 75, 100) or the fuzzy linguistically encoded 5-point Likert-based responses in accordance with some of the most frequently fuzzy linguistic scales considered when 5 labels are modelled (see, for instance, Yeh *et al.* [36], Motawa *et al.* [25] for FSL1' and FSL2', respectively). FLS3' will denote the most usual (balanced) semantic representations of the linguistic hierarchies on the right in Figure 2, FLS4' is partially inspired by the unbalanced semantic representation of 5 points by Herrera *et al.* [18]. FLS1' to FLS4' are displayed in Figure 7.

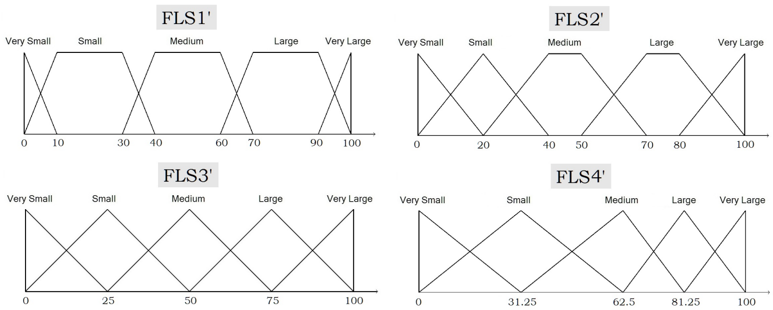



Fig. 7. Examples of four usual fuzzy linguistic scales with 5 terms

The $p$-values in this case are all equal to .000 for any of the five encoded Likert scales, so one can generally conclude that for all the usual significance levels, the FRS-based mean response is significantly different from the encoded Likert-based mean response (when the encoded Likert scale is in {NELikert', FLS1', FLS2', FLS3', FLS4'}). In summary, the mean response is also influenced by the considered scale.

**Remark 4.1** It should be clarified that the imprecise data along the paper have been assumed as coming from an essentially imprecise random element, corresponding to perception, opinion, etc. Eventually, situations can arise for which there might be an underlying real-valued random variable (e.g., the exact relative length w.r.t. the reference line in Case study 2), and we could think in following an 'epistemic' approach, so to draw conclusions about the underlying real-valued random element on the basis of the available imprecise data. However, statistical data analysis in this paper is assumed to be based on the imprecise data but to refer to the imprecise random element supplying them, irrespectively of the fact that imprecise data correspond either to existing data (i.e., an 'ontic' view is considered) or to the imprecise perception of unknown precise data. For a recent detailed discussion about the epistemic/ontic distinction in this setting, see [9]).

## 5 Simulations-based discussion

Simulation studies are to be considered along this section to show whether the conclusions in the preceding one can be generalized, given that to develop general theoretical conclusions is unfeasible in this case. A crucial thought at this stage is that there are not yet suitable realistic models for the distribution of random fuzzy numbers. This makes the simulation process a novel endeavor. On the other hand, in simulating the double data, it should be taken into account that in practice the Likert data are not given independently of FRS ones, and there is a rather systematic reasonable connection between linked data.

In this work simulations of FRS-based data have been inspired by real-life datasets in connection with fuzzy rating scale-based experiments. To generate fuzzy data from a trapezoidal-valued random fuzzy number $\mathcal{X} = \mathrm{Tra}(\inf \mathcal{X}_0, \inf \mathcal{X}_1, \sup \mathcal{X}_1, \sup \mathcal{X}_0)$ Sinova *et al.* [31] suggest to use the characterization, $\mathcal{X} = \mathrm{Tra}\langle X_1, X_2, X_3, X_4\rangle$, where $X_1 = \mathrm{mid}\,\mathcal{X}_1$, $X_2 = \mathrm{spr}\,\mathcal{X}_1 = (\sup \mathcal{X}_1 - \inf \mathcal{X}_1)/2$, $X_3 = \mathrm{lspr}\,\mathcal{X}_0 = \inf \mathcal{X}_1 - \inf \mathcal{X}_0$, $X_4 = \mathrm{uspr}\,\mathcal{X}_0 = \sup \mathcal{X}_0 - \sup \mathcal{X}_1$, that is, $\mathcal{X} = \mathrm{Tra}\langle X_1, X_2, X_3, X_4\rangle = \mathrm{Tra}(X_1 - X_2 - X_3, X_1 - X_2, X_1 + X_2, X_1 + X_2 + X_4)$.

In fact, fuzzy data will be generated by simulating the four real-valued random variables $X_1$, $X_2$, $X_3$ and $X_4$ so that 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 $x_1/x_2$ = center/radius of the core, and $x_3/x_4$ = lower/upper spread of the fuzzy number. To each generated 4-tuple $(x_1, x_2, x_3, x_4)$ we associate the fuzzy number $\text{Tra}\langle x_1, x_2, x_3, x_4 \rangle$.

According to the simulation procedure to be considered, data have been generated from random fuzzy numbers with a bounded reference set and 'mimicking' what we have observed in some real-life examples employing the fuzzy rating scale. Actually, in these examples we have examined the separate behaviour of each of the real-valued components $X_i$, and some convenient models properly fitting such a behaviour.

More concretely, fuzzy data have been generated so that

– 5% 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 to cover three different distributions (namely uniform, an asymmetric one like $p = 1 < 10 = q$, and bell-shaped symmetrical, like $p = q = 5$, see Figure 8). In most of the comparative studies involving simulations, the values from the beta distribution are re-scaled and translated to an interval $[l_0, u_0]$ different from $[0, 1]$, but the bootstrapped two-sample test conclusions are fully irrespective of the re-scaling and translation.

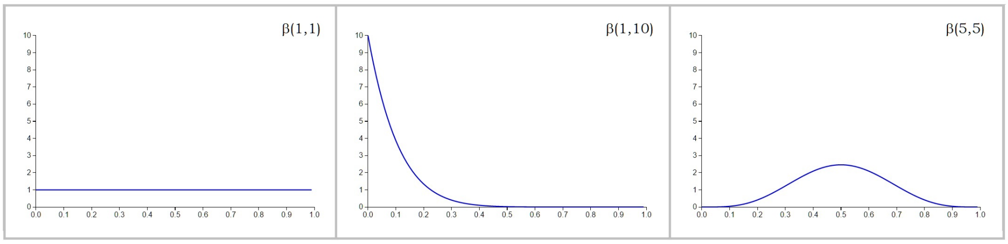



Fig. 8. Density functions of different Beta$(p, q)$ to be used in the simulation studies

– 35% 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:

$$Y_1 \sim \beta(p, q),$$
$$Y_2 \sim \text{Uniform}\Big[0, \min\{1/10, Y_1, 1 - Y_1\}\Big],$$
$$Y_3 \sim \text{Uniform}\Big[0, \min\{1/5, Y_1 - Y_2\}\Big],$$
$$Y_4 \sim \text{Uniform}\Big[0, \min\{1/5, 1 - Y_1 - Y_2\}\Big];$$

– 60% 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:

$$Y_1 \sim \beta(p, q),$$

$$Y_2 \sim \begin{cases} \text{Exp}(200) & \text{if } Y_1 \in [0.25, 0.75] \\ \text{Exp}(100 + 4\,Y_1) & \text{if } Y_1 < 0.25 \\ \text{Exp}(500 - 4\,Y_1) & \text{otherwise} \end{cases}$$

$$Y_3 \sim \begin{cases} \gamma(4, 100) & \text{if } Y_1 - Y_2 \geq 0.25 \\ \gamma(4, 100 + 4\,Y_1) & \text{otherwise} \end{cases}$$

$$Y_4 \sim \begin{cases} \gamma(4, 100) & \text{if } Y_1 + Y_2 \geq 0.25 \\ \gamma(4, 500 - 4\,Y_1) & \text{otherwise.} \end{cases}$$

On the other hand, to mimic the systematic reasonable behaviour in assessing the linked Likert-type data, we will consider the criterion of the minimum $\rho_2$-distance, so that if we assume there are $k = 4$ possible Likert responses, the datum in each of the encoded scales ES $\in$ {NELikert, FLS1, FLS2, FLS3, FLS4, FLS5} (see Figure 5) will be chosen to be the element in the scale showing the lowest $\rho_2$-distance to the FRS datum.

30 samples of size $n$ (with $n \in \{10, 30, 100\}$) have been generated for each of the three considered distributions for $X_1$, namely, $\beta(1,1)$, $\beta(1,10)$ and $\beta(5,5)$. The bootstrapped two-sample test about means for linked samples in Section 3.4 has been now applied (with $B = 1000$), for FRS-based means *vs* ES-based means, where ES $\in$ {NELikert, FLS1, FLS2, FLS3, FLS4, FLS5}. The $p$-values for $n = 10$ and 30 have been gathered in Table 3. Those for $n = 100$ have not been collected since all of them equal .000.

Consequently, one could conclude that for moderate to large sample sizes differences between FRS- and ES-based means are almost generally significant for the usual significance levels. For small sample sizes, this statement sometimes fails, although in many situations differences are also significant.

Analogously, to mimic the systematic reasonable behaviour in assessing the linked Likert-type data when we assume $k = 5$ possible Likert responses, the datum in each of the encoded scales ES $\in$ {NELikert', FLS1', FLS2', FLS3', FLS4'} (see Figure 7) will be chosen to be the element in the scale showing the lowest $\rho_2$ distance to the FRS datum.

30 samples of size $n$ (with $n \in \{10, 30, 100\}$) have been generated for each of the three considered distributions for $X_1$. The bootstrapped two-sample test about means for linked samples in Section 3.4 has been now applied (with $B = 1000$) for FRS-based means *vs* ES-based means, where ES $\in$ {NELikert', FLS1', FLS2', FLS3', FLS4'}. The $p$-values for $n = 10$ and 30 have been gathered in Table 4. Those for $n = 100$ have not been collected since all of them equal .000.

Table 3
Bootstrapped $p$-values of the two-sample test about means (FRS *vs* ES $\in$ {NELikert, FLS1, FLS2, FLS3, FLS4, FLS5}) for simulated linked samples

| $n = 10$ | | | | | | | $n = 30$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_1 \sim \beta(1,1)$ — $k = 4$ | | | | | | | | | | | | | |
| sample \ ES | NELikert | FLS1 | FLS2 | FLS3 | FLS4 | FLS5 | sample \ ES | NELikert | FLS1 | FLS2 | FLS3 | FLS4 | FLS5 |
| S1 | .139 | .020 | .018 | .000 | .004 | .000 | S1’ | .000 | .000 | .006 | .000 | .000 | .000 |
| S2 | .089 | .000 | .011 | .001 | .010 | .000 | S2’ | .002 | .000 | .000 | .000 | .000 | .000 |
| S3 | .131 | .010 | .001 | .009 | .003 | .000 | S3’ | .000 | .000 | .004 | .000 | .001 | .000 |
| S4 | .043 | .041 | .041 | .002 | .049 | .001 | S4’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S5 | .211 | .003 | .001 | .000 | .002 | .000 | S5’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S6 | .044 | .027 | .020 | .002 | .000 | .000 | S6’ | .000 | .000 | .011 | .000 | .001 | .000 |
| S7 | .052 | .010 | .048 | .027 | .012 | .002 | S7’ | .000 | .000 | .001 | .000 | .000 | .000 |
| S8 | .028 | .002 | .035 | .001 | .034 | .000 | S8’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S9 | .215 | .023 | .001 | .007 | .003 | .000 | S9’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S10 | .096 | .032 | .009 | .001 | .031 | .002 | S10’ | .002 | .000 | .000 | .000 | .000 | .000 |
| S11 | .154 | .010 | .015 | .013 | .018 | .013 | S11’ | .003 | .000 | .000 | .000 | .000 | .000 |
| S12 | .088 | .019 | .003 | .004 | .007 | .011 | S12’ | .002 | .001 | .000 | .000 | .000 | .000 |
| S13 | .015 | .014 | .039 | .001 | .008 | .000 | S13’ | .005 | .000 | .000 | .000 | .000 | .000 |
| S14 | .275 | .008 | .001 | .000 | .001 | .001 | S14’ | .000 | .000 | .003 | .000 | .000 | .000 |
| S15 | .031 | .047 | .032 | .014 | .004 | .000 | S15’ | .000 | .000 | .013 | .000 | .000 | .000 |
| S16 | .155 | .010 | .007 | .002 | .001 | .001 | S16’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S17 | .078 | .008 | .012 | .005 | .001 | .001 | S17’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S18 | .045 | .006 | .104 | .022 | .048 | .000 | S18’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S19 | .119 | .029 | .002 | .002 | .015 | .000 | S19’ | .004 | .000 | .000 | .000 | .000 | .000 |
| S20 | .045 | .002 | .017 | .001 | .030 | .000 | S20’ | .000 | .000 | .001 | .000 | .000 | .000 |
| S21 | .071 | .010 | .002 | .000 | .004 | .000 | S21’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S22 | .076 | .005 | .002 | .000 | .003 | .000 | S22’ | .000 | .000 | .005 | .000 | .000 | .000 |
| S23 | .034 | .014 | .163 | .003 | .014 | .000 | S23’ | .001 | .001 | .004 | .000 | .000 | .000 |
| S24 | .091 | .001 | .013 | .002 | .002 | .001 | S24’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S25 | .078 | .008 | .018 | .028 | .005 | .000 | S25’ | .004 | .000 | .000 | .000 | .000 | .000 |
| S26 | .034 | .001 | .016 | .003 | .008 | .000 | S26’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S27 | .040 | .011 | .113 | .010 | .008 | .000 | S27’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S28 | .045 | .000 | .054 | .015 | .018 | .000 | S28’ | .000 | .000 | .001 | .000 | .000 | .000 |
| S29 | .035 | .000 | .004 | .005 | .010 | .001 | S29’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S30 | .073 | .018 | .003 | .003 | .017 | .000 | S30’ | .004 | .000 | .000 | .000 | .000 | .000 |
| $X_1 \sim \beta(1,10)$ — $k = 4$ | | | | | | | | | | | | | |
| sample \ ES | NELikert | FLS1 | FLS2 | FLS3 | FLS4 | FLS5 | sample \ ES | NELikert | FLS1 | FLS2 | FLS3 | FLS4 | FLS5 |
| S1 | .000 | .001 | .000 | .000 | .024 | .000 | S1’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S2 | .121 | .026 | .007 | .006 | .103 | .000 | S2’ | .000 | .000 | .000 | .000 | .002 | .000 |
| S3 | .078 | .074 | .001 | .003 | .265 | .000 | S3’ | .000 | .000 | .000 | .000 | .001 | .000 |
| S4 | .360 | .020 | .003 | .009 | .000 | .002 | S4’ | .000 | .000 | .000 | .000 | .001 | .000 |
| S5 | .025 | .008 | .000 | .002 | .002 | .000 | S5’ | .003 | .000 | .000 | .000 | .000 | .000 |
| S6 | .001 | .014 | .002 | .001 | .067 | .000 | S6’ | .003 | .000 | .000 | .000 | .000 | .000 |
| S7 | .002 | .000 | .001 | .001 | .029 | .000 | S7’ | .010 | .000 | .000 | .000 | .000 | .000 |
| S8 | .057 | .026 | .000 | .003 | .029 | .000 | S8’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S9 | .117 | .016 | .002 | .003 | .249 | .004 | S9’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S10 | .105 | .042 | .007 | .003 | .125 | .000 | S10’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S11 | .009 | .000 | .000 | .000 | .049 | .000 | S11’ | .006 | .000 | .000 | .000 | .000 | .000 |
| S12 | .008 | .001 | .000 | .000 | .067 | .001 | S12’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S13 | .011 | .001 | .011 | .006 | .032 | .000 | S13’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S14 | .137 | .025 | .000 | .001 | .007 | .003 | S14’ | .004 | .000 | .000 | .000 | .000 | .000 |
| S15 | .157 | .021 | .004 | .007 | .077 | .000 | S15’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S16 | .231 | .018 | .004 | .004 | .068 | .000 | S16’ | .004 | .000 | .000 | .000 | .000 | .000 |
| S17 | .068 | .019 | .000 | .000 | .060 | .000 | S17’ | .000 | .000 | .000 | .000 | .001 | .000 |
| S18 | .002 | .000 | .000 | .000 | .030 | .000 | S18’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S19 | .002 | .034 | .004 | .003 | .083 | .005 | S19’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S20 | .115 | .020 | .002 | .000 | .058 | .001 | S20’ | .001 | .000 | .000 | .000 | .001 | .000 |
| S21 | .000 | .001 | .000 | .000 | .073 | .000 | S21’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S22 | .121 | .032 | .006 | .017 | .052 | .000 | S22’ | .005 | .000 | .000 | .000 | .000 | .000 |
| S23 | .276 | .010 | .001 | .003 | .014 | .000 | S23’ | .000 | .000 | .000 | .000 | .002 | .000 |
| S24 | .050 | .061 | .010 | .010 | .538 | .000 | S24’ | .002 | .000 | .000 | .000 | .002 | .000 |
| S25 | .034 | .010 | .000 | .000 | .288 | .000 | S25’ | .000 | .000 | .000 | .000 | .001 | .000 |
| S26 | .137 | .043 | .004 | .004 | .004 | .000 | S26’ | .000 | .000 | .000 | .000 | .001 | .000 |
| S27 | .238 | .013 | .004 | .005 | .014 | .000 | S27’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S28 | .009 | .047 | .002 | .002 | .191 | .000 | S28’ | .004 | .000 | .000 | .000 | .000 | .000 |
| S29 | .006 | .005 | .000 | .000 | .014 | .000 | S29’ | .004 | .000 | .000 | .000 | .000 | .000 |
| S30 | .068 | .000 | .000 | .000 | .008 | .000 | S30’ | .001 | .000 | .000 | .000 | .000 | .000 |
| $X_1 \sim \beta(5,5)$ — $k = 4$ | | | | | | | | | | | | | |
| sample \ ES | NELikert | FLS1 | FLS2 | FLS3 | FLS4 | FLS5 | sample \ ES | NELikert | FLS1 | FLS2 | FLS3 | FLS4 | FLS5 |
| S1 | .084 | .008 | .006 | .004 | .002 | .000 | S1’ | .005 | .000 | .000 | .000 | .000 | .000 |
| S2 | .031 | .006 | .040 | .002 | .015 | .000 | S2’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S3 | .024 | .003 | .034 | .000 | .019 | .000 | S3’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S4 | .032 | .010 | .131 | .011 | .013 | .000 | S4’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S5 | .020 | .003 | .038 | .000 | .003 | .000 | S5’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S6 | .008 | .015 | .027 | .001 | .006 | .000 | S6’ | .002 | .000 | .000 | .000 | .000 | .000 |
| S7 | .195 | .004 | .009 | .002 | .002 | .000 | S7’ | .000 | .000 | .003 | .000 | .000 | .000 |
| S8 | .045 | .004 | .124 | .000 | .005 | .000 | S8’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S9 | .055 | .034 | .087 | .009 | .004 | .000 | S9’ | .004 | .000 | .000 | .000 | .000 | .000 |
| S10 | .009 | .002 | .170 | .010 | .032 | .000 | S10’ | .002 | .000 | .000 | .000 | .000 | .000 |
| S11 | .025 | .002 | .167 | .011 | .018 | .000 | S11’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S12 | .126 | .007 | .003 | .001 | .004 | .000 | S12’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S13 | .015 | .002 | .085 | .002 | .013 | .000 | S13’ | .003 | .000 | .000 | .000 | .000 | .000 |
| S14 | .000 | .001 | .110 | .000 | .078 | .000 | S14’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S15 | .010 | .000 | .022 | .001 | .017 | .000 | S15’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S16 | .018 | .005 | .039 | .008 | .000 | .000 | S16’ | .003 | .000 | .000 | .000 | .000 | .000 |
| S17 | .119 | .002 | .011 | .000 | .004 | .000 | S17’ | .000 | .000 | .003 | .000 | .000 | .000 |
| S18 | .033 | .018 | .009 | .016 | .005 | .000 | S18’ | .000 | .000 | .001 | .000 | .000 | .000 |
| S19 | .004 | .000 | .008 | .017 | .023 | .000 | S19’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S20 | .018 | .018 | .066 | .000 | .015 | .000 | S20’ | .000 | .000 | .003 | .000 | .000 | .000 |
| S21 | .111 | .012 | .025 | .007 | .003 | .000 | S21’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S22 | .019 | .006 | .040 | .000 | .001 | .000 | S22’ | .001 | .000 | .002 | .000 | .000 | .000 |
| S23 | .023 | .004 | .154 | .001 | .027 | .000 | S23’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S24 | .090 | .011 | .029 | .003 | .002 | .000 | S24’ | .001 | .000 | .000 | .000 | .000 | .000 |
| S25 | .113 | .007 | .015 | .002 | .004 | .000 | S25’ | .002 | .000 | .000 | .000 | .000 | .000 |
| S26 | .001 | .002 | .312 | .000 | .122 | .000 | S26’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S27 | .037 | .000 | .007 | .000 | .000 | .000 | S27’ | .000 | .000 | .000 | .000 | .000 | .000 |
| S28 | .011 | .001 | .073 | .023 | .004 | .000 | S28’ | .000 | .000 | .006 | .000 | .000 | .000 |
| S29 | .073 | .037 | .144 | .010 | .005 | .000 | S29’ | .004 | .000 | .000 | .000 | .000 | .000 |
| S30 | .021 | .000 | .015 | .018 | .000 | .000 | S30’ | .001 | .000 | .001 | .000 | .000 | .000 |

Table 4

Bootstrapped $p$-values of the two-sample test about means (FRS *vs* ES$\in${NELikert', FLS1', FLS2', FLS3', FLS4'}) for simulated linked samples

| $n = 10$ | | | | | | $n = 30$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_1 \sim \beta(1,1)$ — $k = 5$ | | | | | | | | | | | |
| sample \ ES | NELikert' | FLS1' | FLS2' | FLS3' | FLS4' | sample \ ES | NELikert' | FLS1' | FLS2' | FLS3' | FLS4' |
| S1 | .076 | .021 | .032 | .010 | .049 | S1' | .000 | .001 | .001 | .000 | .001 |
| S2 | .019 | .008 | .029 | .000 | .005 | S2' | .000 | .001 | .001 | .000 | .000 |
| S3 | .004 | .098 | .025 | .006 | .003 | S3' | .009 | .000 | .000 | .000 | .000 |
| S4 | .018 | .023 | .115 | .077 | .003 | S4' | .001 | .007 | .007 | .004 | .005 |
| S5 | .007 | .044 | .031 | .005 | .089 | S5' | .000 | .009 | .002 | .000 | .002 |
| S6 | .090 | .008 | .003 | .002 | .007 | S6' | .000 | .001 | .001 | .001 | .000 |
| S7 | .016 | .122 | .025 | .054 | .007 | S7' | .001 | .000 | .000 | .000 | .000 |
| S8 | .132 | .037 | .011 | .049 | .002 | S8' | .000 | .000 | .000 | .002 | .000 |
| S9 | .012 | .036 | .014 | .002 | .087 | S9' | .000 | .003 | .002 | .000 | .000 |
| S10 | .050 | .300 | .053 | .006 | .003 | S10' | .000 | .045 | .021 | .024 | .006 |
| S11 | .060 | .017 | .138 | .033 | .037 | S11' | .000 | .022 | .007 | .002 | .001 |
| S12 | .012 | .011 | .014 | .008 | .010 | S12' | .000 | .059 | .004 | .000 | .000 |
| S13 | .049 | .041 | .143 | .044 | .001 | S13' | .000 | .001 | .000 | .000 | .000 |
| S14 | .069 | .170 | .009 | .001 | .007 | S14' | .000 | .002 | .000 | .000 | .000 |
| S15 | .003 | .008 | .055 | .002 | .021 | S15' | .000 | .004 | .000 | .000 | .000 |
| S16 | .009 | .218 | .124 | .020 | .014 | S16' | .000 | .001 | .000 | .000 | .000 |
| S17 | .010 | .010 | .008 | .002 | .000 | S17' | .000 | .000 | .000 | .000 | .000 |
| S18 | .065 | .143 | .043 | .007 | .008 | S18' | .000 | .000 | .000 | .000 | .000 |
| S19 | .000 | .046 | .064 | .014 | .022 | S19' | .000 | .000 | .003 | .001 | .000 |
| S20 | .069 | .027 | .044 | .043 | .000 | S20' | .000 | .009 | .023 | .001 | .000 |
| S21 | .033 | .002 | .044 | .020 | .016 | S21' | .000 | .003 | .004 | .000 | .000 |
| S22 | .016 | .020 | .034 | .030 | .010 | S22' | .000 | .003 | .000 | .000 | .000 |
| S23 | .009 | .005 | .013 | .011 | .021 | S23' | .000 | .004 | .000 | .000 | .000 |
| S24 | .094 | .035 | .021 | .037 | .005 | S24' | .000 | .014 | .002 | .007 | .006 |
| S25 | .008 | .042 | .005 | .002 | .017 | S25' | .000 | .000 | .001 | .000 | .000 |
| S26 | .055 | .006 | .063 | .016 | .000 | S26' | .000 | .003 | .000 | .000 | .000 |
| S27 | .014 | .076 | .018 | .026 | .009 | S27' | .000 | .000 | .000 | .000 | .000 |
| S28 | .051 | .023 | .012 | .015 | .010 | S28' | .000 | .000 | .000 | .001 | .000 |
| S29 | .112 | .057 | .043 | .038 | .006 | S29' | .000 | .000 | .000 | .000 | .000 |
| S30 | .073 | .036 | .010 | .006 | .014 | S30' | .002 | .000 | .007 | .000 | .000 |
| $X_1 \sim \beta(1,10)$ — $k = 5$ | | | | | | | | | | | |
| sample \ ES | NELikert' | FLS1' | FLS2' | FLS3' | FLS4' | sample \ ES | NELikert' | FLS1' | FLS2' | FLS3' | FLS4' |
| S1 | .004 | .039 | .172 | .017 | .009 | S1' | .005 | .032 | .000 | .000 | .000 |
| S2 | .053 | .123 | .035 | .014 | .006 | S2' | .004 | .008 | .000 | .000 | .000 |
| S3 | .096 | .037 | .002 | .000 | .000 | S3' | .000 | .097 | .003 | .000 | .000 |
| S4 | .091 | .019 | .035 | .019 | .012 | S4' | .000 | .018 | .000 | .001 | .000 |
| S5 | .091 | .262 | .097 | .071 | .002 | S5' | .004 | .017 | .000 | .000 | .000 |
| S6 | .049 | .111 | .059 | .003 | .000 | S6' | .001 | .010 | .000 | .000 | .000 |
| S7 | .017 | .256 | .129 | .052 | .025 | S7' | .001 | .012 | .000 | .000 | .000 |
| S8 | .022 | .527 | .121 | .074 | .022 | S8' | .000 | .318 | .003 | .000 | .000 |
| S9 | .024 | .150 | .003 | .000 | .040 | S9' | .003 | .023 | .001 | .000 | .000 |
| S10 | .012 | .071 | .012 | .017 | .013 | S10' | .000 | .000 | .001 | .000 | .000 |
| S11 | .057 | .008 | .018 | .000 | .000 | S11' | .001 | .001 | .001 | .000 | .000 |
| S12 | .090 | .106 | .019 | .048 | .027 | S12' | .000 | .012 | .000 | .000 | .000 |
| S13 | .064 | .127 | .062 | .051 | .008 | S13' | .000 | .000 | .000 | .000 | .000 |
| S14 | .090 | .219 | .105 | .021 | .015 | S14' | .004 | .014 | .000 | .000 | .000 |
| S15 | .095 | .325 | .044 | .042 | .019 | S15' | .000 | .002 | .000 | .000 | .000 |
| S16 | .000 | .000 | .001 | .000 | .000 | S16' | .000 | .001 | .000 | .000 | .000 |
| S17 | .026 | .213 | .043 | .003 | .053 | S17' | .000 | .010 | .001 | .000 | .000 |
| S18 | .042 | .458 | .079 | .127 | .043 | S18' | .000 | .059 | .000 | .000 | .000 |
| S19 | .125 | .089 | .029 | .008 | .002 | S19' | .001 | .001 | .000 | .000 | .000 |
| S20 | .045 | .123 | .007 | .010 | .019 | S20' | .003 | .012 | .000 | .000 | .000 |
| S21 | .009 | .435 | .067 | .034 | .003 | S21' | .002 | .000 | .000 | .000 | .000 |
| S22 | .148 | .091 | .002 | .052 | .021 | S22' | .003 | .016 | .000 | .000 | .000 |
| S23 | .073 | .058 | .084 | .013 | .013 | S23' | .009 | .000 | .000 | .000 | .000 |
| S24 | .119 | .264 | .021 | .023 | .023 | S24' | .002 | .008 | .000 | .000 | .000 |
| S25 | .106 | .014 | .026 | .023 | .004 | S25' | .002 | .000 | .000 | .000 | .000 |
| S26 | .026 | .369 | .104 | .043 | .042 | S26' | .000 | .001 | .000 | .000 | .000 |
| S27 | .052 | .383 | .034 | .051 | .011 | S27' | .012 | .053 | .000 | .000 | .000 |
| S28 | .115 | .075 | .021 | .027 | .029 | S28' | .001 | .005 | .000 | .000 | .000 |
| S29 | .025 | .000 | .001 | .000 | .000 | S29' | .002 | .006 | .000 | .000 | .000 |
| S30 | .011 | .102 | .016 | .022 | .008 | S30' | .011 | .015 | .000 | .000 | .000 |
| $X_1 \sim \beta(5,5)$ — $k = 5$ | | | | | | | | | | | |
| sample \ ES | NELikert' | FLS1' | FLS2' | FLS3' | FLS4' | sample \ ES | NELikert' | FLS1' | FLS2' | FLS3' | FLS4' |
| S1 | .004 | .002 | .008 | .018 | .034 | S1' | .000 | .000 | .000 | .000 | .000 |
| S2 | .029 | .021 | .018 | .008 | .003 | S2' | .000 | .000 | .000 | .000 | .000 |
| S3 | .014 | .063 | .023 | .033 | .025 | S3' | .000 | .000 | .000 | .000 | .000 |
| S4 | .003 | .003 | .005 | .002 | .051 | S4' | .000 | .000 | .000 | .000 | .000 |
| S5 | .026 | .038 | .036 | .013 | .020 | S5' | .000 | .000 | .000 | .000 | .000 |
| S6 | .045 | .032 | .001 | .018 | .024 | S6' | .000 | .000 | .000 | .000 | .000 |
| S7 | .070 | .014 | .014 | .024 | .007 | S7' | .000 | .000 | .000 | .000 | .000 |
| S8 | .029 | .028 | .014 | .017 | .008 | S8' | .000 | .000 | .000 | .000 | .000 |
| S9 | .018 | .006 | .028 | .017 | .006 | S9' | .000 | .001 | .000 | .001 | .000 |
| S10 | .022 | .010 | .012 | .007 | .008 | S10' | .000 | .000 | .000 | .000 | .000 |
| S11 | .077 | .006 | .002 | .004 | .006 | S11' | .000 | .000 | .000 | .000 | .000 |
| S12 | .004 | .006 | .007 | .029 | .043 | S12' | .000 | .000 | .000 | .000 | .000 |
| S13 | .011 | .006 | .007 | .003 | .008 | S13' | .000 | .001 | .001 | .000 | .000 |
| S14 | .017 | .078 | .090 | .043 | .022 | S14' | .000 | .000 | .000 | .000 | .000 |
| S15 | .045 | .038 | .000 | .033 | .011 | S15' | .000 | .000 | .000 | .000 | .000 |
| S16 | .025 | .065 | .037 | .014 | .003 | S16' | .000 | .000 | .000 | .000 | .000 |
| S17 | .078 | .020 | .019 | .020 | .003 | S17' | .000 | .000 | .000 | .000 | .000 |
| S18 | .018 | .015 | .010 | .003 | .001 | S18' | .000 | .001 | .000 | .000 | .000 |
| S19 | .015 | .036 | .014 | .011 | .002 | S19' | .000 | .000 | .000 | .000 | .000 |
| S20 | .011 | .043 | .038 | .020 | .026 | S20' | .000 | .000 | .000 | .000 | .000 |
| S21 | .018 | .031 | .032 | .015 | .006 | S21' | .000 | .000 | .000 | .000 | .000 |
| S22 | .019 | .008 | .011 | .007 | .009 | S22' | .000 | .000 | .000 | .000 | .000 |
| S23 | .002 | .027 | .067 | .002 | .000 | S23' | .000 | .000 | .000 | .000 | .000 |
| S24 | .007 | .002 | .002 | .001 | .019 | S24' | .000 | .000 | .000 | .000 | .000 |
| S25 | .049 | .002 | .003 | .003 | .018 | S25' | .000 | .000 | .000 | .000 | .000 |
| S26 | .065 | .008 | .012 | .004 | .016 | S26' | .000 | .000 | .000 | .000 | .000 |
| S27 | .004 | .096 | .069 | .029 | .001 | S27' | .000 | .000 | .000 | .000 | .000 |
| S28 | .015 | .020 | .029 | .010 | .005 | S28' | .000 | .000 | .000 | .000 | .000 |
| S29 | .011 | .013 | .007 | .004 | .017 | S29' | .000 | .001 | .000 | .000 | .000 |
| S30 | .001 | .011 | .025 | .000 | .001 | S30' | .000 | .000 | .000 | .000 | .000 |

As for $k = 4$, for $k = 5$ one could conclude that for large sample sizes differences between FRS- and ES-based means are almost generally significant for the usual significance levels. For moderate sample sizes, the situation is quite close to that for $k = 4$, but in case of symmetric behavior for $X_1$ differences between FRS- and ES-based means can be eventually significant. For small sample sizes, differences are also significant in many cases, but the significance is definitely less general.

It is not surprising from the simulations collected in Tables 3 and 4 that the effect of the chosen scale of measurement is less intense for small sample sizes, like $n = 10$. If one has 10 possible data in the simulation examples, one can get at most $k \in \{4, 5\}$ different Likert values and 10 different FRS ones for them, whence differences in variability, diversity, and so on, are not that big; when $n$ increases, such differences in variability, diversity, and so on, also increase and lead to more significant differences among means.

## 6 Concluding remarks

The preceding analyses have been performed for other metrics like the generalized metric by Bertoluzza *et al.* [4]. Conclusions are very similar, mainly due to the fact that the considered statistic $\mathsf{T}_n$ (Section 3.4) involves the metric in both the numerator and the denominator. More concretely, when the squared distance between spreads/radius is substantially less weighted than the one between the mid-points/centers, the differences between means for the two involved scales are often slightly less significant when sample sizes are small to rather moderate. For large sample sizes, $p$-values are almost constantly equal to .000.

Such a low influence of the choice of the $L^2$ metric is illustrated in Table 5 by comparing some conclusions in Table 2 in connection with Case study 1 for distances

$$D_\theta(\widetilde{U}, \widetilde{V}) = \sqrt{\int_{[0,1]} \left( [\text{mid}\, \widetilde{U}_\alpha - \text{mid}\, \widetilde{V}_\alpha]^2 + \theta \cdot [\text{spr}\, \widetilde{U}_\alpha - \text{spr}\, \widetilde{V}_\alpha]^2 \right) d\alpha}$$

(Bertoluzza *et al.* [4]) with weights $\theta \in \{1, 1/3, .1\}$ (notice that $D_1 = \rho_2$).

To consider much lower weights $\theta$ would entail almost neglecting imprecision (associated with spreads). Alternatively, if we consider fully ignoring imprecision by defuzzifying the FRS-based values in Case study 1 through their Yager's indicators [35] (coined by Nasibov [26] as the weighted averaging based on levels), $\text{wabl}^{\varphi}(\widetilde{U}) = \int_{[0,1]} \text{mid}\, \widetilde{U}_\alpha \, d\alpha$, we can check that all the involved differences are also highly significant.

Table 5
Bootstrapped $p$-values of the two-sample test about means for linked samples (FRS *vs* encoded scale in {NELikert, FLS1}) for $\rho_2$ and other $L^2$ metrics

| | $\rho_2$ | | $D_{1/3}$ | | $D_{0.1}$ | |
|---|---|---|---|---|---|---|
| item \ $\mathcal{X}'$ | NELikert | FLS1 | NELikert | FLS1 | NELikert | FLS1 |
| $R.1$ | .000 | .000 | .000 | .000 | .010 | .000 |
| $R.2$ | .000 | .000 | .000 | .003 | .000 | .019 |
| $R.3$ | .000 | .016 | .060 | .023 | .000 | .036 |
| $M.1$ | .000 | .000 | .000 | .000 | .038 | .008 |
| $M.2$ | .000 | .002 | .000 | .007 | .001 | .042 |
| $M.3$ | .002 | .000 | .004 | .000 | .001 | .004 |
| $S.1$ | .000 | .000 | .000 | .010 | .018 | .057 |
| $S.2$ | .000 | .000 | .000 | .001 | .001 | .018 |
| $S.3$ | .000 | .000 | .000 | .001 | .004 | .034 |

In summary, we can conclude that in dealing with data from imprecise-valued random magnitudes, statistical conclusions concerning the central tendency would be clearly affected by the scale considered to rate such magnitudes. Since FRS are more informative and diverse than the other two scales, and they capture imprecision in an accurate way, we consider their use should be encouraged in statistical analyses, since statistical conclusions would be also more accurate.

## Acknowledgments

The authors are thankful to the reviewers of the original and Guest Editors of the special issue because of their valuable suggestions and comments. The research in this paper has been partially supported by the Principality of Asturias/FEDER Grant GRUPIN14-101 and the Spanish Ministry of Economy and Competitiveness Grant MTM2015-63971-P. Their financial support is gratefully acknowledged.

## References

[1] R.J. Aumann, Integrals of set-valued functions. J. Math. Anal. Appl. 12 (1965) 1–12.

[2] S. Bajpai, A. Sachdeva, J.P. Gupta, Security risk assessment: Applying the concepts of fuzzy logic. J. Hazard. Mater. 173 (2010) 258–264.

[3] E. Benoit, Expression of uncertainty in fuzzy scales based measurements. Measurement 46 (2013) 3778–3782.

[4] C. Bertoluzza, N. Corral, A. Salas, On a new class of distances between fuzzy numbers. Math & Soft Comput. 2 (1995) 71–84.

[5] A. Blanco-Fernández, M.R. Casals, A. Colubi, N. Corral, M. García-Bárzana, M.A. Gil, G. González-Rodríguez, M.T. López, M.A. Lubiano, M. Montenegro, A.B. Ramos-Guajardo, S. de la Rosa de Sáa, B. Sinova, A distance-based statistical analysis of fuzzy number-valued data. Int. J. Approx. Reas. 55 (2014) 1487–1501.

[6] A. Blanco-Fernández, M.R. Casals, A. Colubi, N. Corral, M. García-Bárzana, M.A. Gil, G. González-Rodríguez, M.T. López, M.A. Lubiano, M. Montenegro, A.B. Ramos-Guajardo, S. de la Rosa de Sáa, B. Sinova, Rejoinder on "A distance-based statistical analysis of fuzzy number-valued data." Int. J. Approx. Reas. 55 (2014) 1601–1605.

[7] J. Cai, Q. Li, L. Li, H. Peng, Y. Yang, A hybrid FCASO-SQP method for solving the economic dispatch problems with valve-point effects. Energy 38 (2012) 346–353.

[8] A. Colubi, G. González-Rodríguez, M.A. Gil, W. Trutschnig, Nonparametric criteria for supervised classification of fuzzy data. Int. J. Approx. Reas. 52 (2011) 1272–1282.

[9] I. Couso, D. Dubois, Statistical reasoning with set-valued information: Ontic vs. epistemic views. Int. J. Approx. Reas. 55 (2014) 1502–1518.

[10] S. de la Rosa de Sáa, M.A. Gil, G. González-Rodríguez, M.T. López, M.A. Lubiano, Fuzzy rating scale-based questionnaires and their statistical analysis. IEEE T. Fuzzy Syst. 23 (2015) 111–126.

[11] P. Diamond, P. Kloeden, Metric spaces of fuzzy sets. Fuzzy Sets Syst. 35 (1990) 241–249.

[12] D. Dubois, H. Prade, Additions of interactive fuzzy numbers. IEEE Trans. Automat. Control 26 (1981) 926–936.

[13] M. Freyd, A graphic rating scale for teachers. The J. Educ. Res. 8(5) (1923) 433–439.

[14] J. Ghneim, Fuzzy concept. In: Encyclopedia of Psychometrics - Wikipedia. URL: http://en.wikipedia.org/wiki/Fuzzy_concept, (October 27, 2014) 61–68.

[15] M.A. 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 (2015) 182–191.

[16] G. González-Rodríguez, A. Colubi, M.A. Gil, Fuzzy data treated as functional data. A one-way ANOVA test approach. Comp. Stat. Data Anal. 56(4) (2012) 943–955.

[17] G. González-Rodríguez, A. Colubi, M.A. Gil, P. D'Urso, An asymptotic two dependent samples test of equality of means of fuzzy random variables. In: Proc. IASC-ERS-COMPSTAT'2006 (2006) 689–695.

[18] F. Herrera, E. Herrera-Viedma, L. Martínez, A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. IEEE T. Fuzzy Syst. 16(2) (2008) 354–370.

[19] B. Hesketh, B. Griffin, V. Loh, A future-oriented retirement transition adjustment framework. J. Vocat. Behav. 79 (2011) 303–314.

[20] 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–Mach. Stud. 29 (1988) 21–35.

[21] R. Likert, A technique for the measurement of attitudes. Arch. Psychol. 22(140) (1932) 1–55.

[22] M.A. Lubiano, S. de la Rosa de Sáa, M. Montenegro, B. Sinova, M.A. Gil, Descriptive analysis of responses to items in questionnaires. Why not using a fuzzy rating scale? Inform. Sci. 360 (2016) 131–148.

[23] M.A. Lubiano, M. Montenegro, B. Sinova, S. de la Rosa de Sáa, M.A. Gil, Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications. Eur. J. Oper. Res. 251 (2016) 918–929.

[24] M.A. Lubiano, A. Salas, M.A. Gil, A hypothesis testing-based discussion on the sensitivity of means of fuzzy data with respect to data shape. Fuzzy Sets Syst. (http://dx.doi.org/10.1016/j.fss.2016.10.015)

[25] 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.

[26] E.N. Nasibov, O. Baskan, A. Mert, A learning algorithm for level sets weights in weighted level-based averaging method. Fuzzy Optim. Dec. Making 4 (2005) 279–291.

[27] A.P. Picon, N.R.S. Ortega, R. Watari, C. Sartor, I.C.N. Sacco, Classification of the severity of diabetic neuropathy: a new approach taking uncertainties into account using fuzzy logic. Clinics 67(2) (2012) 151–156.

[28] M.L. Puri, D.A. Ralescu, The concept of normality for fuzzy random variables. Ann. Probab. 11 (1985) 1373–1379.

[29] M.L. Puri, D.A. Ralescu, Fuzzy random variables. J. Math. Anal. Appl. 114 (1986) 409–422.

[30] J. Sanz, A. Fernández, H. Bustince, F. Herrera, A genetic tuning to improve the performance of Fuzzy Rule-Based Classification Systems with Interval-Valued Fuzzy Sets: Degree of ignorance and lateral position. Int. J. Appr. Reas. 52 (2011) 751–766.

[31] B. Sinova, M.A. 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.

[32] B. Sinova, M.A. Gil, S. van Aelst, $M$-Estimates of location for the robust central tendency of fuzzy data. IEEE T. Fuzzy Syst. 24(4) (2016) 945–956.

[33] W. Trutschnig, M.A. Lubiano, SAFD: Statistical Analysis of Fuzzy Data. https://CRAN.R-project.org/package=SAFD (2015).

[34] R.A. Vitale, $L_p$ metrics for compact, convex sets. J. Approx. Theor. 45 (1985) 280–287.

[35] R.R. Yager, A procedure for ordering fuzzy subsets of the unit interval. Inform. Sci. 24 (1981) 143–161.

[36] 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.

[37] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning. Part 1. Inform. Sci. 8 (1975) 199–249; Part 2. Inform. Sci. 8 (1975) 301–353; Part 3. Inform. Sci. 9 (1975) 43–80.

[38] L.A. Zadeh, Is there a need for fuzzy logic? Inform. Sci. 178 (2008) 2751–2779.