Psicothema
Psicothema 2015, Vol. 27, No. 2, 182-191
doi: 10.7334/psicothema2014.268
ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Analyzing data from a fuzzy rating scale-based questionnaire. A case study

María Ángeles Gil, María Asunción Lubiano, Sara de la Rosa de Sáa and Beatriz Sinova
Universidad de Oviedo

Abstract

Background: The fuzzy rating scale was introduced to cope with the imprecision of human thought and experience in measuring attitudes in many fields of Psychology. The flexibility and expressiveness of this scale allow us to properly describe the answers to many questions involving psychological measurement. **Method:** Analyzing the responses to a fuzzy rating scale-based questionnaire is indeed a critical problem. Nevertheless, over the last years, a methodology is being developed to analyze statistically fuzzy data in such a way that the information they contain is fully exploited. In this paper, a summary review of the main procedures is given. **Results:** The methods are illustrated by their application on the dataset obtained from a case study with nine-year-old children. In this study, children replied to some questions from the well-known TIMSS/PIRLS questionnaire by using a fuzzy rating scale. The form could be filled in either on the computer or by hand. **Conclusions:** The study indicates that the requirements of background and training underlying the fuzzy rating scale are not too demanding. Moreover, it is clearly shown that statistical conclusions substantially often differ depending on the responses being given in accordance with either a Likert scale or a fuzzy rating scale.

Keywords: fuzzy numbers, fuzzy rating scale, Likert scale, Statistics with fuzzy data.

Resumen

Análisis de datos de un cuestionario basado en la escala de valoración fuzzy. Estudio de caso. **Antecedentes:** la escala de valoración difusa se introdujo para abordar la imprecisión inherente al pensamiento humano y la experiencia al medir actitudes en muchos campos de la Psicología. La flexibilidad y expresividad de esta escala permiten describir apropiadamente las respuestas a la mayoría de las cuestiones que involucran mediciones psicológicas. **Método:** analizar las respuestas a cuestionarios basados en dicha escala supone un problema crítico. No obstante, en los últimos años se está desarrollando una metodología específica para el análisis estadístico de datos difusos que explota toda la información disponible. En este trabajo se recoge un resumen de los procedimientos más relevantes. **Resultados:** los métodos se ilustrarán mediante su aplicación a los datos de un estudio realizado con niños de nueve años. En él, los niños han respondido a algunas cuestiones del conocido cuestionario TIMSS/PIRLS recurriendo a un formulario basado en la escala de valoración difusa y en formato impreso o digital. **Conclusiones:** en primer lugar, el estudio muestra que los requisitos previos de formación y entrenamiento para cumplimentar tal formulario son poco exigentes. En segundo lugar, se verifica que a menudo las conclusiones estadísticas difieren sustancialmente dependiendo de que las respuestas se den según escala Likert o de valoración difusa.

Palabras clave: números difusos, escala de valoración difusa, escala Likert, Estadística con datos difusos.

Fuzzy sets were formally introduced fifty years ago by Zadeh (1965) as a tool to provide “a natural way of dealing with the problems in which the source of imprecision is the absence of sharply defined criteria of class membership….” As imprecisely defined classes, fuzzy sets were anticipated to “play an important role in human thinking”, and Psychology was foreseen as one of the promising fields of application.

Ten years later, Zadeh (1975) formalized the concept of *linguistic variable* as a means to address ill-defined phenomena which cannot be properly described in conventional quantitative terms. More concretely, linguistic values were interpreted as labels for ‘fuzzy restrictions’, which associate each value in a referential set with its ‘degree of compatibility’ in [0,1] with such a restriction.

Actually, by looking at the Web of Science, Psychology is one of the subject categories having the most works citing Zadeh’s seminal paper (1965), and many other papers/books have been devoted to applying fuzzy sets to psychological studies (e.g., Zétényi, 1988; Smithson & Oden, 1999; or Stoklasa, Talášek, & Musilová, 2014).

Among the various psychological topics the Fuzzy Set Theory is applied to, one can mention learning disability (e.g., Horvath, 1988), comparison of psychophysical methods (e.g., Garriga Trillo, & Dorn, 1991), psychotherapy (e.g., Horowitz & Malle, 1993), occupational preferences (e.g., Hesketh, Hesketh, Hansen, & Goranson, 1995), false memories (e.g., Leding, 2013), user’s web navigation patterns (e.g., Agarwal & Agarwal, 2005), analysis of questionnaires (e.g., Coppi, Giordani, & D’Urso, 2006), developmental psychology (e.g., Van Dijk & Van Geert, 2009),

---

Received: November 27, 2014 • Accepted: March 6, 2015
Corresponding author: María Ángeles Gil
Departamento de Estadística e I.O. y D.M.
Universidad de Oviedo
33007 Oviedo (Spain)
e-mail: magil@uniovi.es

work adjustment to retirement transition (Hesketh, Griffin, & Loh, 2011), signal detection analysis (Szalma & Hancock, 2013), linguistic prototypes (e.g., Ávila-Muñoz & Sánchez-Sáez, 2014), etc.

Although influential reputed psychologists like Osheron and Smith (1981, 1982) argued by considering specific examples that fuzzy sets cannot be employed to deal properly with concepts and their 'calculus', these arguments have been refuted over the years (see, for instance, Bělohlávek, Klir, Lewis, & Way, 2002, 2009; Bělohlávek & Klir, 2011; Wierman, 2013).

As outlined by Walsh, Teo, and Baydala (2014), "fuzzy logic reflects how people actually think by assigning gradations of meaning… Both the act of measurement and the use of statistical techniques to make sense of numerical observations with absolute certainty are questionable. Fuzzy Logic adds another critical dimension to the historical problems of psychologists' investigative language and uncritical reliance on quantification." The spirit of these assertions has been well captured by the so-called fuzzy rating scale, introduced by Hesketh, Pryor, and Hesketh (1988) in the framework of psychological measurement through questionnaires.

Questionnaires are designed to assess many domains of psychology-related issues, such as perceptions, opinions, emotional states, etc., and the responses to the involved questions have been usually given by means of Likert scales (see, for instance, Peña-Suárez, Muñiz, Campillo-Álvarez, Fonseca-Pedrero, & García Cueto, 2013; Castillo, Tomás, Ntoumanis, Bartholomew, Duda, & Balaguer, 2014, for some recent examples). Different studies have been carried out to discuss the influence of the number of categories/points of the Likert scale on the reliability of the analysis of these responses. They usually coincide in pointing out that increasing the number of categories results in an increase of the variability, information and reliability (see, for instance, Tomás & Oliver, 1998; Lozano, García-Cueto, & Muñiz, 2008).

Actually, to some extent, the ideal situation would be increasing the number of choices to a continuum, but one cannot achieve it by using a natural language, as outlined by Sowa (2013). If one aims to really exploit the individual differences in responding to questionnaires, there is a need for a rich and expressive scale in which "something can be meaningful although we cannot name it" (Ghneim, 2013). All this agrees with Zadeh's attempt to 'precisiate' natural language by treating the associated measurements as objects of computation. In this sense, Zadeh (2008) wisely asserts that "Paradoxically, one of the principal contributions of fuzzy logic... is its high power of 'precisiation' of what is imprecise."

The *fuzzy rating scale* by Hesketh et al. (1988) takes advantages of three of the main skills of fuzzy sets, namely, the abilities to formalize mathematically imprecise valuations, to 'precisiate' them in a continuous way allowing infinite nuances, and to develop mathematical computations with them. In contrast to questionnaires using Likert scales, for which responses to questions are constrained to choosing one within a list of prefixed labels, questionnaires based on the fuzzy rating scale have a free format (i.e., in a sense, it combines the visual analogue and fuzzy linguistic scales). This freedom definitely entails a gain of information and accuracy, which should not be neglected when analyzing responses from a statistical viewpoint.

Aiming to achieve this goal, Hesketh et al., (2011) have indicated the need for statistical techniques (especially inferential ones) for fuzzy data. Although a few studies have been carried out to analyze fuzzy rating scale-based data (e.g., Hesketh et al., 1988, 199; Takemura, 1999, 2007), these studies are descriptive ones, often involving a certain defuzzification process.

Over the last few years, a statistical methodology to analyze fuzzy data is being developed through several studies (see, for instance, Blanco-Fernández et al., 2014, for a recent review). This methodology is based on the so-called random fuzzy numbers, which allow us to develop statistics with fuzzy data within a probabilistic setting and, furthermore, to preserve the key ideas from the real-valued case. The novelties of this methodology with respect to the previous studies in the literature analyzing fuzzy rating scale-based data are that it adds inferential procedures and, instead of involving defuzzification processes, each fuzzy datum is treated as a whole, so no relevant information is lost.

This paper aims to illustrate the application of this statistical methodology to analyze data coming from a fuzzy rating scale-based questionnaire, as well as to corroborate that conclusions from this analysis usually differ from those of the data analysis for the original Likert scale-based questionnaire, both questionnaires having been conducted on the same sample.

Method

Participants

To show how the fuzzy rating scale works by means of a real-life example, sixty-nine 9-year-old children from the fourth grade students of the Colegio San Ignacio in Oviedo (Asturias, Spain) have participated in the study.

The distribution of participants by sex, questionnaire format, mark taken in the last exam and other aspects can be found in the supplementary material http://bellman.ciencias.uniovi.es/SMIRE/FuzzyRatingScaleQuestionnaire.html.

Instrument

In this subsection, we shall firstly recall the main general tools (preliminaries) and later we shall describe the specific ones considered for the case study.

General tools: Firstly, we recall the notions from Fuzzy Set Theory which are required to model and handle the responses from a fuzzy rating scale-based questionnaire. Later, this questionnaire and the scale underlying it are detailed. Thirdly, a summary of the already developed procedures to analyze fuzzy responses is presented.

The responses to questions from a fuzzy rating scale-based questionnaire are assumed to be modelled as fuzzy numbers, where (see Zadeh, 1975; Dubois & Prade, 1979):

Definition 1. A *fuzzy number* is a fuzzy subset of the space of real numbers, that is, a mapping $\tilde{U} : \Re \to [0,1]$ which is normal, convex and has compact levels (i.e., the α-level sets given by

$$\tilde{U}_\alpha = \{x \in \Re : \tilde{U}(x) \geq \alpha\} \text{ if } \alpha \in (0,1),$$
$$\tilde{U}_0 = cl\,\{x \in \Re : \tilde{U}(x) > 0\}$$

with '*cl*' denoting the closure of the corresponding set, are closed and bounded intervals). $\tilde{U}(x)$ can be interpreted as the 'degree of compatibility' of x with the 'property' defining $\tilde{U}$.

Consequently, fuzzy numbers are a 'level-wise' extension of interval values, in such a way that levels add gradualness to the imprecision associated with intervals.

A particular type of fuzzy number which is easy-to-use for both drawing and computations is that of a *trapezoidal fuzzy number*. A trapezoidal fuzzy number $\tilde{U} = Tra(a,b,c,d) = Tra(inf\ \tilde{U}_0, inf\ \tilde{U}_1, sup\ \tilde{U}_1, sup\ \tilde{U}_0)$, with inf = infimum, sup = supremum, can be characterized in terms of its corresponding α-level sets, given by the expression $\tilde{U}_\alpha = [\alpha b + (1-\alpha)a, \alpha c + (1-\alpha)d]$, $\alpha \in [0,1]$, so its meaning can be interpreted as follows:

- Values in $\tilde{U}_1 = [b, c]$ are viewed as fully compatible with $\tilde{U}$.
- Values in $\tilde{U}_0 = [a, d]$, are viewed as compatible with $\tilde{U}$ to some extent.
- Grades of compatibility are linearly interpolated.

Although trapezoidal fuzzy numbers are a special type of fuzzy number, the way the 0- and the 1-level are 'interpolated' would scarcely affect the statistical conclusions.

In handling fuzzy data for statistical purposes, computations often involve arithmetic with fuzzy numbers, more concretely, the sum and the product by scalars. In extending the sum and the product by a scalar from real to fuzzy numbers, the common way is to use Zadeh's extension principle (Zadeh, 1975), which is equivalent to extending level-wise the arithmetic interval.

Definition 2. Let $\tilde{U}$ and $\tilde{V}$ be two fuzzy numbers. The *sum of $\tilde{U}$ and $\tilde{V}$* is the fuzzy number $\tilde{U}+\tilde{V}$ such that for each $\alpha \in [0,1]$

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

Definition 3. Let $\tilde{U}$ be a fuzzy number and let γ a real number. The *product of $\tilde{U}$ by γ* is the fuzzy number $\gamma \cdot \tilde{U}$ such that for each $\alpha \in [0,1]$

$$\left(\gamma \bullet \tilde{U}\right)_\alpha = \begin{cases} \left[\gamma \bullet \inf \tilde{U}_\alpha, \gamma \bullet \sup \tilde{U}_\alpha\right] & if\ \gamma \geq 0, \\ \left[\gamma \bullet \sup \tilde{U}_\alpha, \gamma \bullet \inf \tilde{U}_\alpha\right] & if\ \gamma < 0 \end{cases}$$

Figure 1 shows examples of the sum and the product by scalar related to fuzzy numbers. It should be emphasized that $(\tilde{U}+(-1)\cdot\tilde{U})_\alpha$ is not usually equal to {0} (except in case $\tilde{U}$ reduces to a real number). As a consequence of this, in dealing with fuzzy numbers, one should be very careful and be aware that computations cannot be developed as a simple extension of those for real numbers.

Most of the inconveniences associated with this assertion have been overcome by analyzing fuzzy data through the use of appropriate distances. Among the distances for fuzzy numbers that can be employed (see, for instance, Diamond & Kloeden, 1990), the following ones are outstanding examples whose suitability for statistical purposes has already been shown:

Definition 4. Let $\tilde{U}$ and $\tilde{V}$ be two fuzzy numbers. The *distance between $\tilde{U}$ and $\tilde{V}$* can be measured, among others, by any of the following values:

$$\rho_2\left(\tilde{U},\tilde{V}\right) = \sqrt{\int_0^1 \left\{ \left(\left[\inf \tilde{U}_\alpha - \inf \tilde{V}_\alpha\right]^2 + \left[\sup \tilde{U}_\alpha - \sup \tilde{V}_\alpha\right]^2 \right) / 2 \right\}}\ d\alpha,$$

$$\rho_1\left(\tilde{U},\tilde{V}\right) = \sqrt{\int_0^1 \left\{ \left(\left|\inf \tilde{U}_\alpha - \inf \tilde{V}_\alpha\right| + \left|\sup \tilde{U}_\alpha - \sup \tilde{V}_\alpha\right| \right) / 2 \right\}}\ d\alpha,$$

both introduced by Diamond and Kloeden (1990), and

$$D\left(\tilde{U},\tilde{V}\right) = \sqrt{\int_0^1 \int_0^1 \left[\tilde{U}_\alpha^{[\tau]} - \tilde{V}_\alpha^{[\tau]}\right]^2 d\tau\, d\alpha},$$

introduced by Bertoluzza, Corral, and Salas (1995), with $\tilde{U}_\alpha^{[\tau]} = \tau \cdot \sup \tilde{U}_\alpha + (1\text{-}\tau) \cdot \inf \tilde{U}_\alpha$.

Whereas the first and third distances correspond to the so-called L^2 type (often considered in studies concerning means), the second distance is of L^1 type (interesting in connection with some robust location measures).

The *fuzzy rating scale* has been introduced (Hesketh et al., 1988) as an approach allowing us to combine a free-response format with a fuzzy valuation. In the fuzzy rating scale, along a continuous line between two end-points:

- A respondent selects or draws a 'representative position/interval' of her/his rating (i.e., the 1-level = set of points which she/he considers to be fully compatible with such a rating).

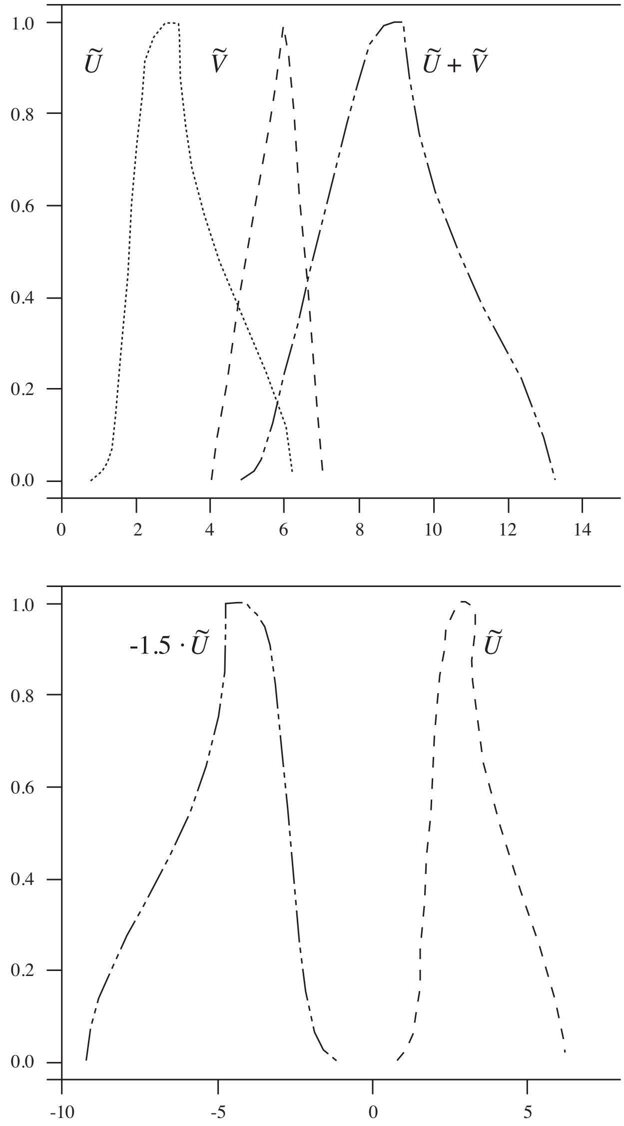


Figure 1. *Examples of the sum (on the top) and the product by a scalar (on the bottom) of fuzzy numbers*

- And he/she also indicates ‘latitudes of acceptance’ on either side by determining the highest and lowest possible positions for the respondent rating (i.e., the 0-level = set of points which she/he considers to be compatible to some extent with such a rating).

Once these two steps are given (see Figure 2 on the top), a trapezoidal fuzzy number is drawn by using linear interpolation (see Figure 2 on the bottom).

The free response format allows respondents to use a continuum of ‘values’ which, in general, cannot be universally ordered. As it has been empirically shown by De la Rosa de Sáa, Gil, González-Rodríguez, López, and Lubiano (2015), the use of the expressive fuzzy rating scale leads, in most cases, to more accurate statistical conclusions, as one explores and exploits more information and variability than with the use of the Likert ones.

When responses from a questionnaire are given in accordance with a qualitative/categorical scale, the statistical analysis is usually based on the unique quantifiable involved aspects, namely, frequencies of the few modalities/categories and maybe positions.

When responses are given in accordance with the fuzzy rating scale, they add some other quantifiable aspects: certain intervals (the level sets) and certain gradualness (the degree of compatibility with the response). Consequently, the statistical analysis of these responses should not neglect this quantitative information, and it should also exploit the variability inherited from the use of the free response format as much as possible.

Over the past years a methodology is being developed aiming to extend some of the valuable procedures in dealing with real-valued data to analyze fuzzy ones. A crucial role in this methodology is played by the model for the random mechanism generating fuzzy number-valued data (random fuzzy number), which is well-stated within the probabilistic setting and enables preserving most of the key concepts in statistics with real-valued data (e.g., the bias of an estimator, the *p*-value of a test, etc.).

Definition 5. Given a random experiment (modelled by means of a probability space (Ω, A, P), a *random fuzzy number* (RFN for short) associated with it is a fuzzy number-valued mapping χ defined on Ω and such that for all $\alpha \in [0,1]$ the real-valued mappings $\inf \chi_\alpha$ and $\sup \chi_\alpha$ (with $\inf \chi_\alpha(\omega) = \inf(\chi(\omega))_\alpha$ and $\sup \chi_\alpha(\omega) = \sup(\chi(\omega))_\alpha$ for all $\omega \in \Omega$) are real-valued random variables.

Random fuzzy numbers (introduced in a more general framework by Puri and Ralescu, 1986) can be formalized in equivalent ways (see, for instance, Blanco-Fernández et al., 2014) guaranteeing that one can refer properly to the (induced) distribution of a random fuzzy number, the independence of random fuzzy numbers, and so on.

The analysis of fuzzy number-valued data often concerns some summary measures, like:

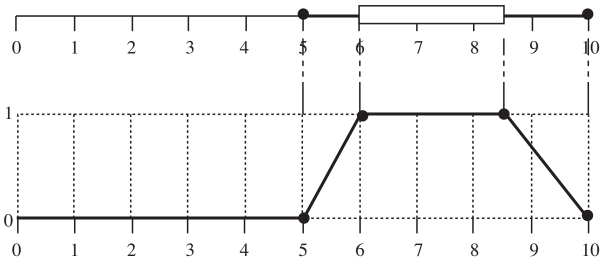


Figure 2. *Example of a fuzzy rating scale-based response*

- (Fuzzy-valued) central tendency measures (e.g., the Aumann-type mean value and the 1-norm median of an RFN).
- And (real-valued) dispersion measures (e.g., the variance).

These measures are now recalled.

Definition 6. Given a random experiment and an associated RFN χ, the (Aumann-type) *mean value* of χ (Puri & Ralescu, 1986) is the fuzzy number $\tilde{E}(\chi)$, if it exists, such that for all $\alpha \in [0,1]$

$$(\tilde{E}(\chi))_\alpha = [E(\inf \chi_\alpha), E(\sup \chi_\alpha)]$$

(with *E* denoting the mean value of the associated random variable). In particular, if $\tilde{x}^{(n)} = (\tilde{x}_1, \ldots, \tilde{x}_n)$ is a sample of observations from the RFN χ, then the corresponding *sample mean* is given by the fuzzy number $\overline{\tilde{x}^{(n)}}$ such that for all $\alpha \in [0,1]$

$$\overline{\tilde{x}^{(n)}} = \left[\frac{\inf(\tilde{x}_1)_\alpha + \ldots + \inf(\tilde{x}_n)_\alpha}{n}, \frac{\sup(\tilde{x}_1)_\alpha + \ldots + \sup(\tilde{x}_n)_\alpha}{n}\right]$$

Definition 7. Given a random experiment and an associated RFN χ, the (1-norm) median of χ (Sinova, Gil, Colubi, & Van Aelst, 2012) is the fuzzy number $\widetilde{M}(\chi)$ such that for all $\alpha \in [0,1]$

$$(\widetilde{M}(\chi))_\alpha = [Me(\inf \chi_\alpha), Me(\sup \chi_\alpha)]$$

(with *Me* denoting the median of the associated random variable, and using the convention of the mid-point of the class of medians in case of non-uniqueness). In particular, if $\tilde{x}^{(n)} = (\tilde{x}_1, \ldots, \tilde{x}_n)$ is a sample of observations from the RFN χ, then the corresponding *sample median* is given by the fuzzy number $\widetilde{M}(\tilde{x}^{(n)})$ such that for all $\alpha \in [0,1]$.

$$\widetilde{M}(\tilde{x}^{(n)}) = [Me\{\inf(\tilde{x}_1)_\alpha, \ldots, \inf(\tilde{x}_n)_\alpha\}, Me\{\sup(\tilde{x}_1)_\alpha, \ldots, \sup(\tilde{x}_n)_\alpha\}]$$

Consequently, the computation of these two central tendency measures reduces to that of the measures they extend on certain real-valued random variables.

Definition 8. Given a random experiment and an associated RFN χ, the ρ_2 *-variance* (respectively, *D-variance*) of χ (Lubiano, Gil, López-Díaz, & López, 2000) is the real number $Var_{\rho_2}(\chi)$ (respectively, $Var_D(\chi)$), if it exists, given by

$$Var_{\rho_2}(\chi) = E([\rho_2(\chi, \tilde{E}(\chi))]^2)$$
$$(\text{respectively, } Var_D(\chi) = E([D(\chi, \tilde{E}(\chi))]^2)).$$

The mean value and the median of an RFN, as defined above, extend those for the real-valued case and preserve the usual properties held in that case.

In developing a methodology for the statistical analysis of fuzzy number-valued data associated with a random experiment, in addition to the above-recalled summary measures, one should be aware, regarding inferential purposes, that

- ‘Realistic’ families of models (like normal, Poisson, etc., in the real-valued case) for distributions of RFNs have not yet been established.
- And there are no Central Limit Theorems for RFNs that are directly applicable.

Thanks to the use of appropriate distances between fuzzy data, the concept of random fuzzy numbers, and the existence of a bootstrapped Central Limit Theorem for generalized random elements (Giné & Zinn, 1990), the preceding shortcomings can be overcome and large samples are not usually required.

Among the inferential problems and procedures which have already been tackled:

- Some studies have been carried out concerning the estimation of fuzzy- and real-valued 'parameters' of the distribution of an RFN on the basis of the information provided by a sample of independent observations from it (cf., Lubiano & Gil, 1999; Sinova et al., 2012).
- Some bootstrap techniques have been developed to *test statistical hypotheses*, among others, *about the fuzzy-valued means* of RFNs by considering 'two-sided' hypotheses expressed in terms of the involved distances; the *one-sample test* has been discussed by considering a bootstrap approach in Montenegro, Colubi, Casals, and Gil (2004) and González-Rodríguez, Montenegro, Colubi, and Gil (2006); the *two-sample test* has been discussed for independent samples in Montenegro, Casals, Lubiano, and Gil (2001) and for dependent samples in González-Rodríguez, Colubi, Gil, and D'Urso (2006); the *k-sample case* (one-way ANOVA test) for independent/dependent samples have been discussed in González-Rodríguez, Colubi, and Gil (2012) and in Montenegro, López-García, Lubiano, and González-Rodríguez (2009), respectively; the *factorial ANOVA* has been examined in Nakama, Colubi, and Lubiano (2010); and a test for the *equality of variances* can be found in Ramos-Guajardo, and Lubiano (2012).

Although this statistical methodology is often based on results which are stated in a rather abstract setting, the application of the methodology is definitely much simpler. Furthermore, the R package SAFD (Statistical Analysis of Fuzzy Data) by Trutschnig and Lubiano (http://cran.r-project.org/web/packages/SAFD/index.html) implements most of the computations involved in these procedures.

Specific tools: In 2011, the TIMSS (Trends in International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study) have joined to provide countries with the opportunity to assess their fourth grade students in three fundamental curricular areas: mathematics, science, and reading.

In collaboration with the Spanish Institute of Educational Evaluation (INEE) a data analysis has been developed (see Corral-Blanco, Zurbano-Fernández, Blanco-Fernández, García-Honrado, & Ramos-Guajardo, 2013) on data collected through some of the TIMSS/PIRLS questionnaires conducted in Spanish schools. These questionnaires are standard ones, and most of the responses have to be chosen among those on a Likert scale with 4 points, namely, *Strongly Disagree* (**A1**), *Somewhat Disagree* (**A2**), *Somewhat Agree* (**A3**) and *Strongly Agree* (**A4**).

To show how the fuzzy rating scale works by means of a real-life example, the conducted questionnaire was designed on the basis of the selection of a few questions from the Student questionnaire TIMSS/PIRLS 2011 (see http://timss.bc.edu/timss2011/downloads/T11_StuQ4.pdf). The chosen nine questions were based on the 4-point Likert scale **A1-A4** in the original TIMSS/PIRLS questionnaire, but for this experiment, they were formulated with a double-type response (namely, the **A1–A4** and the fuzzy rating scale-based with reference set the interval [0,10], see Figure 3). So, these nine questions were not modified with respect to the original, but simply the second way of responding was added.

The questionnaire was designed in both paper-and-pencil and computerized format (see http://bellman.ciencias.uniovi.es/SMIRE/FuzzyRatingScaleQuestionnaire.html for both versions, the second one in Spanish).

Procedure

Teachers decided that 24 of the students should fill out the paper-and-pencil and 45 of them should complete the computerized version.

The training of the fourth grade students in the study was carried out by providing them with some instructions to fill out the questionnaire. As it is difficult to explain how to draw a trapezoidal fuzzy set to students at this level, because they do not have the required background about real-valued functions yet, we made use of the notion of trapezium. No remarkable problems were found either in the training, which lasted up to 15 minutes, or in the obtained responses being coherent and plausible.

Data analysis

The Likert responses, or the corresponding encoded Likert ones, were analyzed with SPSS, whereas the fuzzy-valued responses were analyzed with R package SAFD.

Two types of studies were carried out, namely,

- The estimates of the median of the 4-point Likert scale-based responses, the estimates of the mean and variance of their 'equidistant' encoding to 0–10 (i.e., **A1** ≡ 0, **A2** ≡ 10/3, **A3** ≡ 20/3, and **A4** ≡ 10), ELikert, and the extended versions of the fuzzy rating scale-based ones, FRS, were computed.
- The equality of means of two or more subpopulations/levels for the ELikert responses, as well as for the FRS ones, was tested.

Results

This section presents the details for the study which has been selected among those performed. It corresponds to the double-responses in connection with reading, math and science.

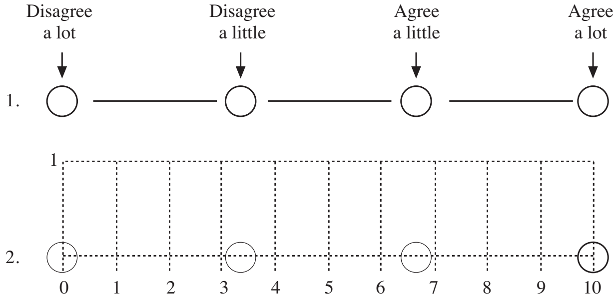


Figure 3. *Example of the double response (4-point Likert and fuzzy rating scale-based question) form*

First analysis (estimates of central tendency and dispersion)

The (estimates of the) mean, median and variance of the double responses (Likert/ELikert and FRS) to the nine questions concerning the three areas are presented in Table 1. The estimates for the Likert/ELikert responses were obtained by well-known computations. In the case of trapezoidal fuzzy data, computations for the Aumann-type value become really simple. To illustrate this assertion, if ($\tilde{r}_{1,1}$, …, $\tilde{r}_{1,69}$) is the sample of responses to Question R.1, with $\tilde{r}_{1,i} = Tra(a_i, b_i, c_i, d_i)$ the corresponding *sample mean* is given by the trapezoidal fuzzy number given in Figure 4 and involve the following computations

$$\frac{1}{69}\sum_{i=1}^{69}\tilde{r}_{1,i} = Tra\left(\frac{1}{69}\sum_{i=1}^{69}\alpha_i, \frac{1}{69}\sum_{i=1}^{69}b_i, \frac{1}{69}\sum_{i=1}^{69}c_i, \frac{1}{69}\sum_{i=1}^{69}d_i\right)$$

The computation of the mean for non-trapezoidal fuzzy data is mostly more complex. Moreover, the median and variance are not that easy to compute, irrespectively of the involved fuzzy data, as they should first be performed for each of the levels, so they cannot be easily illustrated and the use of computer technology is virtually essential.

Second analysis (testing the equality of some means)

The first test to be considered analyzes the influence of the factor ‘mark in the scale 0-10 taken in the last exam’ on the mean response to a question. This factor was assumed in the analysis to act at the levels given by intervals $G1 = [0,6]$, $G2 = [6,8]$, $G3 = [8,9]$, and $G4 = [9,10]$,

The respondents to the nine questions concerning reading, math and science are distributed as shown in Table 2, which indicates that such an asymmetric grouping leads to a rather balanced situation. Although the rows in Table 2 could often coincide for the three questions in each area, this was not the case, as some students did not provide either their marks or their response to some questions.

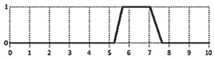
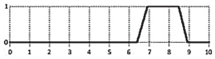
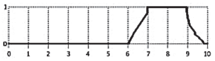
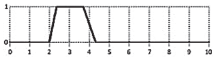
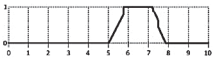
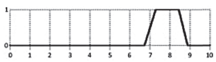
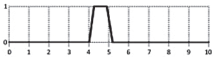
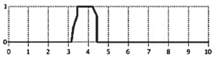
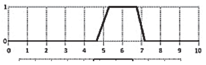
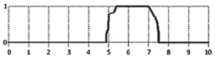
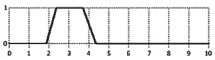
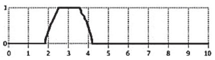
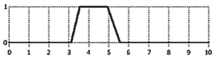
To examine the influence of the factor ‘mark in the scale 0-10 taken in the last exam’ (the factor acting at 4 possible levels, which are assumed to be given by intervals $G1$ - $G4$) on the response to each of the 3 questions posed per area, conclusions will be drawn on the basis of the responses in the considered sample of 69 students. The one-way ANOVA test by González-Rodríguez et al. (2012) was performed to obtain the approximate *p*-values shown in Table 3, together with those for the Kruskal-Wallis test (KW) on the Likert responses.

The test for the FRS case indicates that the marks taken affect more (i.e., the *p*-values are lower for) questions associated with Science than with Math, and those with Math more than those with Reading.

Multiple two-sample independent comparisons were performed by using the Mann-Whitney-Wilcoxon (MWW) test for the Likert responses and the procedure in Montenegro et al. (2001) for the FRS ones (Tables 4 and 5).

Table 1

Sample mean, median and variance of the FRS, Likert and ELikert response along the 9 different questions

	ELikert mean	FRS Aumann-type mean	Likert median	FRS 1-norm median	ELikert variance	FRS D-variance	# Valid responses
R.1	6.3738		**A3**		6.2191	4.7650	*68*
R.2	8.2099		**A4**		4.8236	3.1600	*67*
R.3	2.1885		**A1**		10.2874	8.2858	*67*
M.1	6.5672		**A3**		9.4243	7.0894	*67*
M.2	8.3341		**A4**		6.2381	5.2719	*66*
M.3	5.8935		**A3**		16.3918	12.2326	*69*
S.1	6.2572		**A3**		9.9034	6.6063	*65*
S.2	2.6553		**A2**		7.1199	5.1205	*64*
S.3	3.9392		**A2**		12.2823	8.2269	*66*

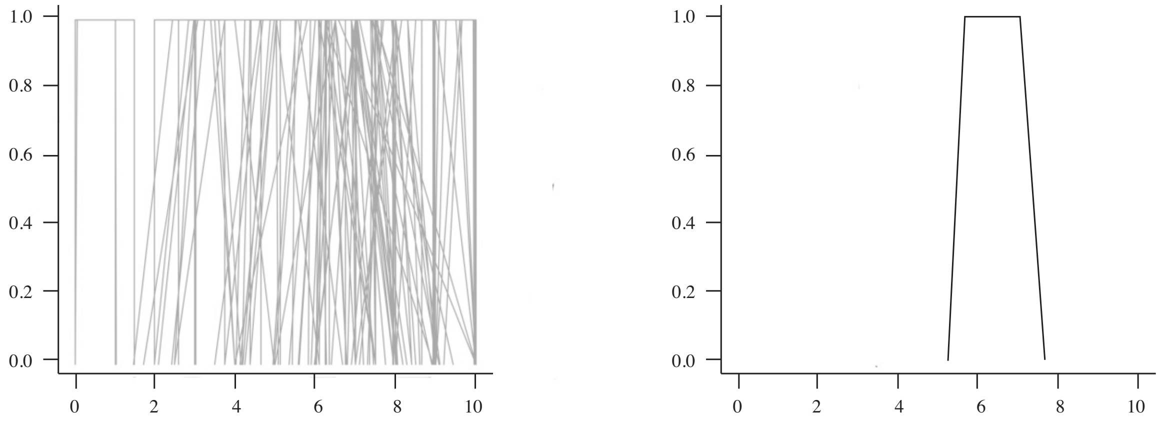


Figure 4. *Responses to Question R.1 and the corresponding Aumann-type mean. On the left, the 69 sample FRS responses $\tilde{r}_{1,i}$. On the right, the sample Aumann-type mean*

Table 2

Absolute frequency distribution of the number of respondents to each of the 9 questions in accordance with the group associated with their marks in the area

		Absolute frequencies			
		G1	**G2**	**G3**	**G4**
READING					
R.1	I like to read things that make me think	5	26	14	7
R.2	I learn a lot from reading	5	26	13	7
R.3	Reading is harder for me than any other subject	5	25	14	7
MATHS					
M.1	I like math	7	18	16	20
M.2	My teacher is easy to understand	7	18	16	19
M.3	Math is harder for me than any other subject	7	19	16	20
SCIENCE					
S.1	My teacher taught me to discover science in daily life	9	23	10	16
S.2	I read about science in my spare time	8	22	10	17
S.3	Science is harder for me than any other subject	10	23	10	17

Table 3

p-values in testing the equality of mean responses for different levels of the mark taken in the last exam

QUESTION	**KW Likert p-value**	**FRS p-value**
R.1	**.026***	.084
R.2	**.012***	**.000*****
R.3	**.045***	.100
M.1	**.005****	**.000*****
M.2	.167	**.000*****
M.3	.462	.067
S.1	**.008****	**.001****
S.2	.457	**.030***
S.3	**.006****	**.001****

* *p*<.05, ** *p*<.01, *** *p*<.001

Some additional studies with data in this case study were performed and collected in http://bellman.ciencias.uniovi.es/SMIRE/FuzzyRatingScaleQuestionnaire.html (supplementary material).

Table 4

p-values in the pairwise MWW comparisons of the Likert responses for levels of the mark achieved in the last exam

	p-value					
	G1⇔G2	**G1⇔G3**	**G1⇔G4**	**G2⇔G3**	**G2⇔G4**	**G3⇔G4**
R.1	.115	.107	.073	.747	**.043***	.094
R.2	.071	.443	**.018***	.187	.090	**.024***
R.3	.516	.298	.106	.239	**.023***	.197
M.1	.198	.055	**.011***	.075	**.004****	.626
M.2	.141	.308	.063	.695	.538	.350
M.3	.651	.308	.341	.317	.270	.937
S.1	**.013***	**.013***	.388	.343	.079	**.041***
S.2	.185	.203	.406	.952	.604	.639
S.3	.363	**.002****	**.031***	**.010***	.126	.264

* *p*<.05, ** *p*<.01, *** *p*<.001

Table 5

p-values in the pairwise comparisons of the fuzzy means for levels of the mark taken in the last exam

	p-value					
	G1⇔G2	**G1⇔G3**	**G1⇔G4**	**G2⇔G3**	**G2⇔G4**	**G3⇔G4**
R.1	.091	.059	.144	.300	.512	.766
R.2	.106	.242	**.047***	.427	**.000*****	**.011***
R.3	.625	.217	.414	.078	.506	.821
M.1	.169	**.046***	**.002****	.203	**.010***	.222
M.2	**.010***	.063	**.001****	.578	.225	.138
M.3	.323	.208	.558	.667	.512	.351
S.1	**.007****	**.008****	.557	.283	.073	**.024***
S.2	.090	.145	.612	.915	.204	.308
S.3	.372	**.003****	.059	**.012***	.194	.236

* *p*<.05, ** *p*<.01, *** *p*<.001

Conclusions

The analysis summarized in Table 2 clearly shows *how conclusions differ between using Likert and using FRS data*, and how the latter often allows for a more visible distinction between outputs. In this way, we note substantial differences, such as:

- The 'central tendency' of the response to the same question may be distant from one scale to the other (e.g., those for M.3).
- The central tendency response to two questions can be very close with respect to the Likert/ELikert data, but substantially different with respect to the FRS data (e.g., those for M.3 and S.1).
- The last two columns concerning variance corroborate what has been empirically asserted by De la Rosa de Sáa et al. (2015): although the FRS incorporates a much larger diversity of values, the mean squared deviation is substantially reduced when passing from Likert/ELikert to FRS, so the locations are more representative for FRS data.

From Tables 3, 4 and 5, one can easily conclude that *statistical testing results also differ according to the considered scale*. To achieve a more accurate idea about the essential difference in the information exploited/explored with both scales, we can detail it for responses to M.2 analyzed in Table 6.

Although one cannot make general comparative assertions, from Tables 3 to 5, one can observe that if differences are definitely significant with both scales ($p<.05$), the FRS data seems to reveal these differences more clearly.

Also from Table 5, one can deduce, in connection with the six questions showing lowest p-values for the fuzzy responses that: for Question R.2, main significant differences are between $G4$ and the others; for M.1 and M.2 main differences correspond to those between $G1$ and $G4$; for S.1, S.2 and S.3, the clearest differences are those shown between $G1$ and $G3$. Additional conclusions can be drawn by examining the p-values.

The study in this paper has served not only to show the potential of the fuzzy rating scale and the associated methodology, but it also confirms that, although FRS questionnaires not being as immediate to fill as Likert ones, the required training and background to use FRS questionnaires are not really deep. Therefore, it is an especially advisable scale when one aims to have more accurate and informative conclusions.

It would be interesting as a future research direction to discuss the statistical reliability of the new scale in contrast to the Likert scale ones or even of their fuzzy linguistic conversion.

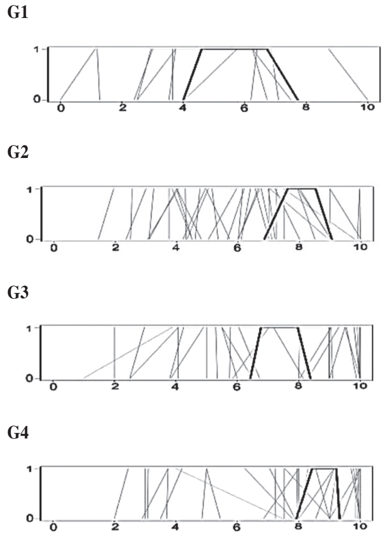
Acknowledgements

This paper is dedicated as a tribute to the father of Fuzzy Logic, Professor Lotfi A. Zadeh, on the occasion of the 50th anniversary of the publication of his seminal paper.

Table 6

Available information and ANOVA p-values in the comparisons of the fuzzy means for levels of the mark taken in the last exam

	Likert-type available information					**Kruskal-Wallis *p*-value**
		A1	**A2**	**A3**	**A4**	
M.2	**G1**	0	2	3	2	.167
	G2	0	1	6	11	
	G3	2	0	5	9	
	G4	0	1	4	14	

	Fuzzy rating-type available information	**FRS-based *p*-value**
M.2		**.000*****

Authors are very grateful to the reviewers of the original manuscript, because of their insightful comments and suggestions to improve the paper. Authors are also grateful to the Colegio San Ignacio in Oviedo-Asturias (Spain), to its staff as well as to its fourth grade students and the parents for their support and understanding. They have allowed us to collect the data in the case study. Actually, the 9-10-year-old students in this school have captured quite properly the spirit and the mechanics behind fuzzy rating scale-based questionnaires. Authors also want to acknowledge their colleague Professor Carlos Carleos because of designing and implementing the computerized format several students have used to answer the questionnaire.

The research in this paper has been partially supported by/ benefited from Principality of Asturias Grants GRUPIN14-101, Research Contract E-33-2015-0040746 (this one for Sinova) and Severo Ochoa BP12012 (this one for De la Rosa de Sáa), and the Spanish Ministry of Economy and Competitiveness Grant MTM2013-44212-P. Their financial support is gratefully acknowledged.

References

Agarwal, S., & Agarwal, P. (2005). A Fuzzy Logic approach to search results' personalization by tracking user's web navigation pattern and Psychology. *In Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence-ICTAI'05* (pp. 318-325).

Ávila-Muñoz, A.M., & Sánchez-Sáez, J.M. (2014). Fuzzy sets and prototype theory representational model of cognitive community structures based on lexical availability trials. *Review of Cognitive Linguistics*, *12*, 133-159.

Bělohlávek, R., & Klir, G.J. (Eds.) (2011). *Concepts and Fuzzy Logic*. Cambridge, USA: The MIT Press.

Bělohlávek, R., Klir, G.J., Lewis III, H.W., & Way, E. (2002). On the capability of Fuzzy Set Theory to represent concepts. *International Journal of General Systems*, *31*, 569-585.

Bělohlávek, R., Klir, G.J., Lewis III, H.W., & Way, E. (2009). Concepts and fuzzy sets: Misunderstandings, misconceptions, and oversights. *International Journal of Approximate Reasoning*, *51*, 23-34.

Bertoluzza, C., Corral, N., & Salas, A. (1995). On a new class of distances between fuzzy numbers. *Mathware & Soft Computing*, *2*, 71-84.

Blanco-Fernández, A., Casals, M.R., Colubi, A., Corral, N., García-Bárzana, M., Gil, M.A., González-Rodríguez, G., López, M.T., Lubiano, M.A., Montenegro, M., Ramos-Guajardo, A.B., De la Rosa de Sáa, S., & Sinova, B. (2014). A distance-based statistical analysis of fuzzy number-valued data. *International Journal of Approximate Reasoning*, *55*, 1487-1501. Rejoinder. *International Journal of Approximate Reasoning*, *55*, 1601-1605.

Castillo, I., Tomás, I., Ntoumanis, N., Bartholomew, K., Duda, J.L., & Balaguer, I. (2014). Psychometric properties of the Spanish version of the Controlling Coach Behaviors Scale in the sport context. *Psicothema*, *26*, 409-414.

Coppi, R., Giordani, P., & D'Urso, P. (2006). Component models for fuzzy data. *Psychometrika*, *71*, 733-761.

Corral-Blanco, N., Zurbano-Fernández, E., Blanco-Fernández, A., García-Honrado, I., & Ramos-Guajardo, A.B. (2013). Structure of the family educational environment: Its influence on performance and differential performance. In *PIRLS-TIMSS 2011 International Study on Progress in Reading Comprehension, Mathematics and Sciences IEA. Volume II. Spanish Report. Secondary Analysis*. Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa (pp. 9-31).

De la Rosa de Sáa, S., Gil, M.A., González-Rodríguez, G., López, M.T., & Lubiano, M.A. (2015). Fuzzy rating scale-based questionnaires and their statistical analysis. *IEEE Transactions on Fuzzy Systems*, *23*, 111-126.

Diamond, P., & Kloeden, P. (1990). Metric spaces of fuzzy sets. *Fuzzy Sets and Systems*, *35*, 241-249.

Dubois, D., & Prade, H. (1979). Fuzzy real algebra: Some results. *Fuzzy Sets and Systems*, *2*, 327-348.

Garriga Trillo, A.J., & Dorn, T. (1991). Medición de la borrosidad: modalidades cruzadas. *Psicothema*, *3*, 423-432.

Giné, E., & Zinn, J. (1990). Bootstrapping general empirical measures. *Annals of Probability*, *18*, 851-869.

González-Rodríguez, G., Colubi, A., & Gil, M.A. (2012). Fuzzy data treated as functional data. A one-way ANOVA test approach. *Computational Statistics & Data Analysis*, *56*, 943-955.

González-Rodríguez, G., Montenegro, M., Colubi, A., & Gil, M.A. (2006). Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data. *Fuzzy Sets and Systems*, *157***,** 2608-2613.

Ghneim, J. (Ed.) (2013). Fuzzy concept. In *Encyclopedia of Psychometrics*, pp. 61-68 - *Wikipedia*. October 27, 2014, URL: http://en.wikipedia.org/wiki/Fuzzy_concept.

Hesketh, B., Griffin, B., & Loh, V. (2011). A future-oriented retirement transition adjustment framework. *Journal of Vocational Behavior*, *79*, 303-314.

Hesketh, B., Hesketh, T., Hansen, J.-I., & Goranson, D. (1995). Use of fuzzy variables in developing new scales from the strong interest inventory. *Journal of Counseling Psychology*, *42*, 85-99.

Hesketh, T., & Hesketh, B. (1994). Computerized fuzzy ratings: The concept of a fuzzy class. *Behavior Research Methods, Instruments & Computers*, *26*, 272-277.

Hesketh, T., Pryor, R., & Hesketh, B. (1988). An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences. *International Journal of Man-Machine Studies*, *29*, 21-35.

Horowitz, L.M., & Malle, B.M. (1993). Fuzzy concepts in psychotherapy research. *Psychotherapy Research*, *3*, 131-148.

Horvath, J.M. (1988). A fuzzy set model of learning disability: identification from clinical data. In T. Zétényi (Ed.), *Fuzzy Sets in Psychology*. Series: Advances in Psychology, vol. 56 (pp. 345-382). Amsterdam: North-Holland, Elsevier.

Leding, J.K. (2013). Need for cognition is related to the rejection (but not the acceptance) of false memories. *American Journal of Psychology*, *126*, 1-10.

Lozano, L.M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, *4*, 73-79.

Lubiano, M.A., & Gil, M.A. (1999). Estimating the expected value of fuzzy random variables in random samplings from finite populations. *Statistical Papers*, *40*, 277-295.

Lubiano, M.A., Gil, M.A., López-Díaz, M., & López, M.T. (2000). The lambda-mean squared dispersion associated with a fuzzy random variable. *Fuzzy Sets and Systems*, *111*, 307-317.

Montenegro, M., Casals, M.R., Lubiano, M.A., & Gil, M.A. (2001). Two-sample hypothesis tests of means of a fuzzy random variable. *Information Sciences*, *133*, 89-100.

Montenegro, M., Colubi, A., Casals, M.R., & Gil, M.A. (2004). Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika*, *59*, 31-49.

Montenegro, M., López-García, M.T., Lubiano, M.A., & González-Rodríguez, G. (2009). A dependent multi-sample test for fuzzy means. In *Abstracts 2nd Workshop ERCIM WG Computing & Statistics* (p. 102).

Nakama, T., Colubi, A., & Lubiano, M.A. (2010). Factorial analysis of variance for fuzzy data. In *Abstracts 3rd Workshop ERCIM WG Computing & Statistics* (p. 88).

Osherson, D.N., & Smith, E.E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*, 35-58.

Osherson, D.N., & Smith, E.E. (1982). Gradedness and conceptual combination. *Cognition*, *12*, 299-318.

Peña-Suárez, E., Muñiz, J., Campillo-Álvarez, Á., Fonseca-Pedrero, E., & García-Cueto, E. (2013). Assessing organizational climate: Psychometric properties of the CLIOR Scale. *Psicothema*, *25*, 137-144.

Puri, M.L., & Ralescu, D.A. (1986). Fuzzy random variables. *Journal of Mathematical Analysis and Applications*, *114*, 409-422.

Ramos-Guajardo, A.B., & Lubiano, M.A. (2012). *K*-sample tests for equality of variances of random fuzzy sets. *Computational Statistics and Data Analysis*, *56*, 956-966.

Sinova, B., Gil, M.A., Colubi, A., & Van Aelst, S. (2012). The median of a random fuzzy number. The 1-norm distance approach. *Fuzzy Sets and Systems, 200*, 99-115.

Smithson, M., & Oden, G.C. (1999). Fuzzy set theory and applications in Psychology. In H.-J. Zimmermann (Ed.), *Practical Applications of Fuzzy Technologies*. Series: The Handbooks of Fuzzy Sets Series Vol. 6 (pp. 557-585). Heidelberg: Springer.

Sowa, J.F. (2013). What is the source of fuzziness? In R. Seising, E. Trillas, C. Moraga & S. Termini (Eds.), *On Fuzziness. A Homage to Lotfi A. Zadeh-Volume 2* (pp. 645-652). Heidelberg: Springer.

Stoklasa, J., Talášek, T., & Musilová, J. (2014). Fuzzy approach - a new chapter in the methodology of Psychology? *Human Affairs*, *24*, 189-203.

Szalma, J.L., & Hancock, P.A. (2013). A signal improvement to signal detection analysis: Fuzzy sdt on the ROCs. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1741-1762.

Takemura, K. (1999). A fuzzy linear regression analysis for fuzzy input-output data using the least squares method under linear constraints and its application to fuzzy rating data. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *3*, 36-41.

Takemura, K. (2007). Ambiguous comparative judgment: fuzzy set model and data analysis. *Japanese Psychological Research*, *49*, 148-156.

Tomás, J.M., & Oliver, A. (1988). Efectos de formato de respuesta y método de estimación en análisis factorial confirmatorio. *Psicothema*, *10*, 197-208.

Van Dijk, M., & Van Geert, P. (2009). The application of Fuzzy Logic principles to Developmental Psychology. In R.E. Vargas (Ed.), *Fuzzy Logic: Theory, Programming and Applications* (pp. 329-336). New York: Nova Sciences Publishers.

Walsh, R.T.G., Teo, T., & Baydala, A. (2014). *A Critical History and Philosophy of Psychology*. Cambridge UK: Cambridge University Press.

Wierman, M.J. (2013). Psychologists: Are they logically fuzzy? In *Joint IFSA World Conference and NAFIPS Annual Meeting* (pp. 854-859).

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

Zadeh, L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning. Part 1. *Information Sciences*, *8*, 199-249; Part 2. *Information Sciences*, *8*, 301-353; Part 3. *Information Sciences*, *9*, 43-80.

Zadeh, L.A. (2008). Is there a need for fuzzy logic? *Information Sciences*, *178*, 2751-2779.

Zétényi, T. (Ed.) (1988). *Fuzzy Sets in Psychology*. Series: Advances in Psychology, Vol. 56. Amsterdam: North-Holland, Elsevier.