

Applying a fuzzy questionnaire in a peer review process

Tamás Jónás, Zsuzsanna Eszter Tóth & Gábor Árva

To cite this article: Tamás Jónás, Zsuzsanna Eszter Tóth & Gábor Árva (2018) Applying a fuzzy questionnaire in a peer review process, Total Quality Management & Business Excellence, 29:9-10, 1228-1245, DOI: [10.1080/14783363.2018.1487616](https://doi.org/10.1080/14783363.2018.1487616)

To link to this article: <https://doi.org/10.1080/14783363.2018.1487616>



Published online: 13 Jun 2018.



Submit your article to this journal [↗](#)



Article views: 28



View Crossmark data [↗](#)

Applying a fuzzy questionnaire in a peer review process

Tamás Jónás^{a*}, Zsuzsanna Eszter Tóth^a and Gábor Árva^b

^a*Institute of Business Economics, Eötvös Loránd University, Budapest, Hungary;* ^b*Department of Management and Business Economics, Budapest University of Technology and Economics, Budapest, Hungary*

This paper introduces the application and the results of a fuzzy number based questionnaire used for evaluating lecturers' performance in a newly launched peer review programme in order to enhance the reliability of peer evaluations. In our approach, the membership function of the fuzzy number is composed of an increasing and a decreasing sigmoid membership function conjuncted by Dombi's intersection operator. Compared to the traditional Likert scale-based evaluation, the proposed fuzzy number based methodology allows peer reviewers to express their uncertainty and the variability of the reviewed lecturer's performance within a single lecture and during a semester in a quantitative way. By utilising the principles of Dombi's Pliant Arithmetic, fuzzy Likert scale based evaluations given by peer reviewers can be cumulated by aggregating the parameters of the left- and right-hand sides of the fuzzy numbers separately. This property results in a much simpler statistical analysis of the gathered data than the methods previously proposed in the literature. A demonstrative case study illustrates our main contributions in peer evaluation.

Keywords: Likert scale; fuzzy rating scale; pliant arithmetic; peer review; higher education; service quality

1. Introduction

How can we assess the quality of educational services offered by a higher education (HE) institution? How can we know reliably whether the expectations of customers and other stakeholders, primarily of students are met? In order to answer these questions, there is a need to design suitable and reliable methodologies to evaluate performance and to identify appropriate measure units to highlight the achieved service performance level (Lupo, 2013; Battisti, Nicolini, & Salini, 2005, 2010).

Measuring service quality and the satisfaction of stakeholders in HE is mainly realised through the application of Likert scales (see e.g. Brochado, 2009; Teeroovengadum, Kama-lanabhan, & Seebaluck, 2016; Nadiri, Kandampully, & Hussain, 2009; Lalla, Facchinetti, & Mastroleo, 2005). Chen (2001) argues that the use of these scales reduces the human perception on one concrete number. Others have also claimed that individuals cannot use an exact number to express their opinion about a given situation. Instead of that, linguistic assessment is preferred to represent that specific numerical value (Herrera & Herrera-Viedma, 2000; Herrera, López, Mendana, & Rodríguez, 1999; Kacprzyk, 1986; Andayani, Hartati, Wardoyo, & Mardapi, 2017; Carrasco, Villar, Hornos, & Herrera-Viedma, 2011; Chang & Wang, 2016; Cabrerizo, López-Gijón, Martínez, Morente-Molinera, & Herrera-Viedma, 2017). In order to consider human perceptions, the fuzzy set theory is increasingly applied in these situations as they improve successfully the reliability of service process measurements and evaluations (Li, 2013; Lupo, 2013; Lin, 2010a; Deng, 2008).

*Corresponding author. Email: jonas@gti.elte.hu

This paper focuses on a challenging problem which is related to how to handle properly the inherent uncertainty of human perceptions. The current study aims to demonstrate the application of fuzzy Likert scales for evaluating lecturers' performance in the peer review programme launched at the Faculty of Economic and Social Sciences. The proposed approach can help to deal with the vagueness arising either from the uncertainty of the reviewers or from the fluctuation of the observed lecturers' performance. By providing a fuzzy Likert scale to evaluate lecturers' performance, reviewers can express their uncertainty, their contrasting perceptions and the variability of the observed lecturers' performance in a quantitative way.

Despite the fact that fuzzy numbers are able to model human judgement more precisely, the lack of simple methods to carry out statistical analyses has been turned out to be an obstacle in their utilisation. Several methods proposed in the literature (see e.g. Frühwirth-Schnatter, 1992 or Amini & Jochem, 2011) require both expertise in fuzzy set theory and high calculation efforts, which could hinder their utilisation. By following Dombi's Pliant Inequality Model (Dombi, 2009), the aggregate evaluation can be computed in a convenient way which can serve the purposes of statistical analysis as well and allows the draw of more reliable managerial conclusions.

The paper is structured as follows. The next section highlights the advantages of fuzzy Likert scales compared to traditional Likert scales and provides our problem statement in the light of the peer review programme available at our faculty. Then the fuzzy set theory based methodology is presented. Finally, a case study demonstrates the advantages of the use of fuzzy Likert scales in peer evaluations. At the end we sum up the main findings and outline future research directions.

2. Literature review: traditional rating vs. fuzzy rating

The evaluation of educational activities may be carried out in several ways and shows great diversity. Berk (2005) gives an overview of the possible varieties of evaluation applied in educational context in case of which commonly employed scales are Likert ones (Hartley, 2014; Murray, 2013). It is a discrete scale by choosing the most appropriate 'values' within a class according to the raters' judgement, opinion, valuation (Gil & González-Rodríguez, 2012) and leads to ordinal data from a set of pre-fixed categories. When applying these Likert-type questionnaires and rating scales, responses are averaged in order to express the mean of a specific teaching performance attribute (De Witte & Rogge, 2011) which is then utilised as a kind of index and applied as a part of formative and/or summative evaluation. Another possible approach is summing the ratings and expressing them as a percentage to the maximal attainable overall rating (Liaw & Goh, 2003). It is also possible to ask students, peer reviewers or other stakeholders to rate the overall, semester-long performance of the lecturer on a single scale (Ellis, Burke, Lomire, & McCormack, 2003).

Rating items in a questionnaire can be considered as a complex task as raters make multiple decisions under uncertainty. The number of 'values' to choose from is small (Gil & González-Rodríguez, 2012) which means that the variability, diversity and subjectivity associated with an accurate rating is usually lost. Another disadvantage originates from the fact that when values are encoded by their relative position in accordance with a certain ranking, differences between codes cannot be interpreted as differences in their magnitude. It means that only statistical conclusions addressed to ordinal data can be reliable and relevant information can be lost (Lubiano, de Sáa, Montenegro, Sinova, & Gil, 2016). Another major issue related to the application of Likert scales is the weighting of aspects as they are usually not equally important for the raters. An additional concern

arising with expressing overall evaluations is the fact that raters' attitudes towards the rated item are not homogeneous as time goes on (see e.g. Tóth Zs, Surman, & Árvá, 2017b). Third, if raters' preferences are heterogeneous, it matters how and to what extent it influences the overall evaluation of a rated item. Average scores are supposed to hide the real situation, namely, the performance of the rated item (Kuzmanovic, Savic, Popovic, & Martic, 2013). Moreover, when Likert-type data are analysed for statistical purposes, the techniques to analyse them are quite limited (Lubiano et al., 2016). Different studies have been carried out to discuss the reliability of the analysis of these responses pointing out that increasing the number of responses results in an increase of information and reliability (Lozano, García-Cueto, & Muñiz, 2008; de Sáa, Gil, González-Rodríguez, López, & Lubiano, 2015). However, it cannot be achieved by using a natural language (Sowa, 2013). To manage these disadvantages there is an alternate approach which takes into account that the nature of most attributes related to evaluations, judgements involve subjectivity and certain imprecision (Lubiano *et al.*, 2016; Quirós, Alonso, & Pancho, 2016).

Hesketh, Pryor, Gleitzman, and Hesketh (1988) proposed the fuzzy rating scale without raters being constrained to choose among a few pre-specified categories. It is expressive enough to find a value in it fitting appropriately the valuation, opinion, judgement involving subjective perceptions in most real life situations (Gil, Lubiano, De Sáa, & Sinova, 2015). This kind of scale has the ability to model the imprecision of human rating evaluations, formalise them mathematically, to 'precisiate' them in a continuous way, and to develop mathematical computation with them (Gil et al., 2015; Calcagni & Lombardi, 2014; Gil & González-Rodríguez, 2012). This approach leads to a fuzzy-valued response format enabling a level of variability and accuracy which would not be captured when using a Likert scale.

In the service quality literature several initiatives confirm the recent shift towards the utilisation of fuzzy ratings (Lin, 2010b). Liou and Chen (2006) demonstrate that fuzzy linguistic assessment of service quality is much closer to human thinking than methods based on crispy numbers. The recent literature provides a number of modifications of the widely applied service quality models (Liu *et al.*, 2015; Mashhadiabdol, Sajadi, & Talebi, 2014; Lupo, 2016; Zhang, Lin, & Ren, 2010).

Fuzzy rating scales have been intensively applied recently in HE context as well to measure quality related issues. Başaran, Kalaycı, and Atay (2011), Lalla et al. (2005) propose a method to implement fuzzy logic in student evaluations of teaching performance. Büyüközkan, Ruan, and Feyzioğlu (2007) evaluate the quality of e-learning websites using fuzzy logic. Yu, Tsang, and Chen (2016) develop a fuzzy linguistic scale to establish a satisfaction index in case of e-learning systems. Lupo (2013) introduces a fuzzy SERVQUAL based method for reliable measurements of education quality. Rouyendegh and Erkan (2013) apply a fuzzy logic based method for academic staff selection, Hammed (2011) introduces one for a student evaluation system.

The membership functions of fuzzy sets can have various shapes (see Figure 1). The aforementioned literatures in fuzzy rating of service quality mainly employ triangular or trapezoidal membership functions, however, they are not always consistent with human thinking and judgement, since in these cases the membership functions have the same slope on the whole interval (see e.g. Hammed, 2011). In reality, the judgement of the respondent changes only slightly around points that represent the worst and the best possible judgement on the service quality, and the same property holds for the evaluation which is most likely to express the rater's judgement. Hammed (2011) also pointed out the disadvantages of triangular membership functions and proposed a Gaussian one to enhance the reliability of

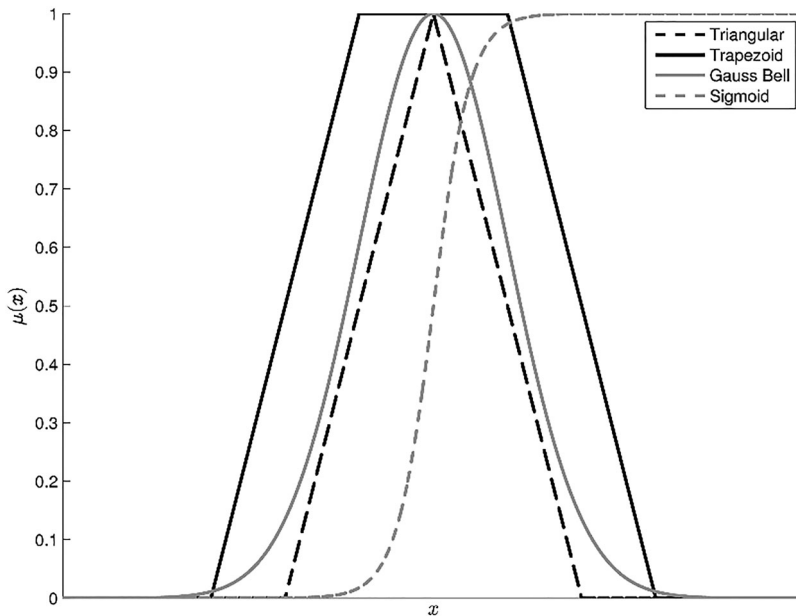


Figure 1. Triangular, trapezoid, Gaussian and sigmoid membership functions.

student evaluations. In our paper, we work with fuzzy numbers the membership function of which is composed of an increasing and a decreasing sigmoid membership function based on its favourable property that it increases or decreases around the above mentioned points less sharply (Dombi, 2008, 2009). That is, the slope of the sigmoid function is not constant, therefore, applying sigmoid functions to establish fuzzy numbers results in a more precise reflection of human thinking and judgement.

By following Dombi's Pliant Inequality Model (Dombi, 2009), the aggregate evaluation based on different raters' judgement can be computed in a convenient way since the parameters of our membership functions are aggregated according to the arithmetic and harmonic mean. This can serve the purposes of statistical analysis as well and allows the draw of more reliable managerial conclusions.

2.1. Peer review framework at the Budapest University of Technology and Economics

Based on the relevant literature and taking international practices into consideration (Samson & McCrea, 2008; Ihsan et al., 2012; Washer, 2006; Blackmore, 2005; Courneya, Pratt, & Collins, 2008; Brent & Felder, 2004), a semester-long peer review of teaching process has been launched at our Faculty. The questionnaires applied to evaluate lecturers' performance include the observation of lectures, the observation of midterm tests and/or exams, and the overall evaluation of the lecturers' semester-long performance in case of which most of the aspects of the lecturer's performance are evaluated using a traditional 5-point Likert scale (Tóth Zs, Andor, & Árvai, 2017a). Peer evaluations include both formative and summative components, since the applied questionnaires consist of two main parts: a numerical scale assessment of specific skills and of performance attributes and a second part standing for obligatory narrative comments.

For the purposes of this study, we focus on the questionnaire applied during classroom teaching observation. In this questionnaire the following performance dimensions have been evaluated on a traditional Likert scale: (1) *Volume, intonation*; (2) *Grammar, intelligibility, speech rate*; (3) *Learning, explaining the terminology*; (4) *Explanatory capability*; (5) *Maintaining students' attention*; (6) *Choosing proper presentation techniques*; (7) *Quality of the slide show*; (8) *Consonance of the slide show with verbal communication*; (9) *Logical structure of the lecture and presentation*. This classroom questionnaire includes a second part as well allowing answers for elaborative questions.

The first results of the peer review programme have been already presented by Tóth Zs, Andor, and Árva (2017a). During the evaluation process, we have come across several major constraints arising from the application of traditional Likert scales. Three main problems of feedbacks have been found that can affect the reliability of results originating from the process of peer reviewing, namely, the uncertainty inherent among evaluations, the variation of lecturers' performance during the semester and the lack of methods available to compare different reviewers' narrative evaluations (Tóth Zs, Surman, and Árva, 2017b). These issues addressed the research problem, namely, the enhancement of the reliability of measurements and evaluations executed in HE context by utilising fuzzy logic.

The evaluations conducted by either peer reviewers, students or other stakeholders can be considered as a complex combination of verbal statements. Evaluators have personal subjective preference or judgment depending on their individual knowledge or experiences. It is suitable for evaluators to use linguistic variables to express their subjective opinions in the assessment process. Researchers usually turn originally verbal data into numerical data by using Likert-type scale questionnaires when they are trying to get the opinions, judgments or valuations of the students, peer reviewers or other stakeholders (Başaran, Kalaycı, and Atay, 2011).

2.2. Drawbacks of peer evaluation

The aspects and evaluation dimensions of the questionnaire based on a traditional Likert scale are not able to cover the fluctuation of the observed lecturers' performance within the same lecture, since evaluations are given at the end of the lecture reflecting an average evaluation in each dimension. Moreover, the retrospective statistics including the mean, the range and the standard deviation of the overall evaluations given at the end of the semester are more likely to reflect the differences between the various reviewers' judgement than the variability of the lecturers' performance.

Narrative comments in the questionnaires are of high importance and kindly welcomed from the reviewees' points of view. They can reflect the variability of the observed lecturer's performance as well, but there is a lack of simple methods to deal with linguistic feedbacks. This leads to an insufficient elaboration of this kind of information gained during the evaluation process. It means that by comparing and evaluating different lecturers' performance based solely on numerical assessment, a remarkable part of the information gained is either lost or is not taken into account, which cannot serve the purposes of the faculty as a whole.

A fuzzy Likert scale can help to overcome the aforementioned difficulties. A fuzzy number depicts not only the differences between different colleagues' judgements, but also the contrasting perceptions and impressions of the same reviewer. The more the uncertainty associated with the judgement and the more unbalanced the observed lecturer's performance are, the more spread out the fuzzy number is.

3. Methodology

3.1. Fuzzy numbers as intersections of two soft inequalities

In our approach, the values on a Likert scale are represented by fuzzy numbers; that is, instead of expressing an opinion by selecting a particular x crisp value on the scale, we allow the evaluator to select an ‘approximately x ’ value that is given by a fuzzy number. We will use sigmoid functions to compose the membership functions of fuzzy numbers.

DEFINITION 1 The sigmoid function $\sigma_a^{(\lambda)}(x)$ with parameter a and λ is given by

$$\sigma_a^{(\lambda)}(x) = \frac{1}{1 + e^{-\lambda(x-a)}}$$

where $x, a, \lambda \in \mathbb{R}$ and λ is nonzero.

The main properties of the sigmoid function $\sigma_a^{(\lambda)}(x)$ are as follows.

- **Range.** The range of $\sigma_a^{(\lambda)}(x)$ is the interval $(0, 1)$
- **Continuity.** $\sigma_a^{(\lambda)}(x)$ is continuous in \mathbb{R}
- **Monotony.**
 - If $\lambda > 0$, then $\sigma_a^{(\lambda)}(x)$ is strictly monotonously increasing
 - If $\lambda < 0$, then $\sigma_a^{(\lambda)}(x)$ is strictly monotonously decreasing
- **Limits.**

$$\lim_{x \rightarrow +\infty} \sigma_a^{(\lambda)}(x) = \begin{cases} 1, & \text{if } \lambda > 0 \\ 0, & \text{if } \lambda < 0 \end{cases}$$

$$\lim_{x \rightarrow -\infty} \sigma_a^{(\lambda)}(x) = \begin{cases} 1, & \text{if } \lambda < 0 \\ 0, & \text{if } \lambda > 0 \end{cases}$$

- **Role of parameters.**

- Parameter a is the locus at which $\sigma_a^{(\lambda)}(x)$ has the value 0.5
- The slope of $\sigma_a^{(\lambda)}(x)$ at a is $\lambda/4$; that is, the parameter λ determines the gradient of function curve at a

Figure 2 shows examples of sigmoid function graphs. Following Dombi’s Pliant Inequality Model (Dombi, 2009), we will use the following definitions where the indexes l and r stand for ‘left’ and ‘right’, respectively, and will be used to denote left hand side and right hand side components of fuzzy numbers.

DEFINITION 2 The soft inequality $\{a_l <_{(\lambda_l)} x\}$ is given by the sigmoid function

$$\{a_l <_{(\lambda_l)} x\} = \sigma_{a_l}^{(\lambda_l)}(x) = \frac{1}{1 + e^{-\lambda_l(x-a_l)}}$$

where $a_l, \lambda_l \in \mathbb{R}, \lambda_l > 0$.

The soft inequality $\{a_l <_{(\lambda_l)} x\}$ represents the truth of the inequality $a_l < x$. This truth value is in the interval $(0, 1)$ and it depends on λ_l . Since λ_l is positive, the greater the variable x is, the higher the truth of the inequality $a_l < x$ is. Notice that the parameter λ_l is responsible for the ‘sharpness’ of the soft inequality $\{a_l <_{(\lambda_l)} x\}$. It can be shown that if

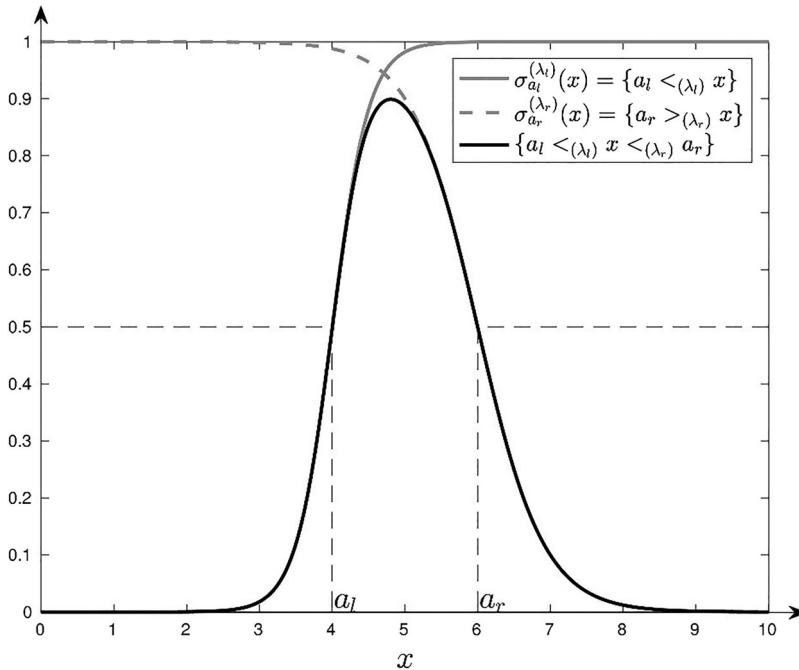


Figure 2. Intersection of two fuzzy sets given by an increasing and a decreasing sigmoid membership function.

$\lambda_l \rightarrow \infty$, then $\{a_l <_{(\lambda_l)} x\} \rightarrow 1$; that is, the truth of the relation $a_l < x$ tends to 1, which means that $a_l < x$ stands in the traditional manner. This result tells us that the soft inequality $\{a_l <_{(\lambda_l)} x\}$ is a generalisation of the crisp inequality $a_l < x$.

DEFINITION 3 *The soft inequality $\{a_r >_{(\lambda_r)} x\}$ is given by the sigmoid function*

$$\{a_r >_{(\lambda_r)} x\} = \sigma_{a_r}^{(\lambda_r)}(x) = \frac{1}{1 + e^{-\lambda_r(x-a_r)}},$$

where $a_r, \lambda_r \in \mathbb{R}, \lambda_r < 0$.

The soft inequalities given by Definitions 1 and 2 allow us to express how much we consider a particular value of x being greater than a_l or less than a_r . The introduced soft inequalities represent fuzzy sets with the membership functions $\sigma_{a_l}^{(\lambda_l)}(x)$ and $\sigma_{a_r}^{(\lambda_r)}(x)$. That is, for any x , $\sigma_{a_l}^{(\lambda_l)}(x)$ and $\sigma_{a_r}^{(\lambda_r)}(x)$ are the membership values of x in the fuzzy sets which contain the numbers that are greater than a_l and less than a_r , respectively. Intersection of these two fuzzy sets represents the soft interval $\{a_l <_{(\lambda_l)} x <_{(\lambda_r)} a_r\}$. We will use the Dombi conjunction operator (Dombi intersection) to implement intersection of two fuzzy sets (Dombi, 2008).

DEFINITION 4 *The Dombi intersection of the fuzzy sets A_1 and A_2 that are given by the membership functions $\mu_{A_1}(x)$ and $\mu_{A_2}(x)$, respectively, is the fuzzy set with membership*

function $\mu_{A_1 \cap A_2}(x)$:

$$\begin{aligned} \mu_{A_1 \cap A_2}(x) &= \mu_{A_1}(x) *_{(D)} \mu_{A_2}(x) \\ &= \frac{1}{1 + \left(\left((1 - \mu_{A_1}(x)) / \mu_{A_1}(x) \right)^\alpha + \left((1 - \mu_{A_2}(x)) / \mu_{A_2}(x) \right)^\alpha \right)^{1/\alpha}} \end{aligned}$$

where $\mu_{A_1}(x), \mu_{A_2}(x) \in (0, 1), \alpha \in \mathbb{R}, \alpha > 0$ and $*_{(D)}$ denotes the Dombi intersection operator.

If we apply the Dombi intersection to $\sigma_{a_l}^{(\lambda_l)}(x)$ and $\sigma_{a_r}^{(\lambda_r)}(x)$ with $\alpha = 1$, we get

$$\begin{aligned} \{a_l <_{(\lambda_l)} x <_{(\lambda_r)} a_r\} &= \sigma_{a_l}^{(\lambda_l)}(x) *_{(D)} \sigma_{a_r}^{(\lambda_r)}(x) \\ &= \frac{1}{1 + (1 - \sigma_{a_l}^{(\lambda_l)}(x)) / \sigma_{a_l}^{(\lambda_l)}(x) + (1 - \sigma_{a_r}^{(\lambda_r)}(x)) / \sigma_{a_r}^{(\lambda_r)}(x)}. \end{aligned}$$

Utilising Definition 2 and Definition 3:

$$\{a_l <_{(\lambda_l)} x <_{(\lambda_r)} a_r\} = \sigma_{a_l}^{(\lambda_l)}(x) *_{(D)} \sigma_{a_r}^{(\lambda_r)}(x) = \frac{1}{1 + e^{-\lambda_l(x-a_l)} + e^{-\lambda_r(x-a_r)}}.$$

Figure 2 shows the intersection of two fuzzy sets given by an increasing and a decreasing sigmoid membership function. The following theorem demonstrates a key property of the soft inequalities given by sigmoid functions.

THEOREM 1 *If A_1, A_2, \dots, A_n are fuzzy sets with the membership functions $\sigma_{a_1}^{(\lambda_1)}(x), \sigma_{a_2}^{(\lambda_2)}(x), \dots, \sigma_{a_n}^{(\lambda_n)}(x)$, respectively, $\text{sgn}(\lambda_1) = \text{sgn}(\lambda_2) = \dots = \text{sgn}(\lambda_n)$, and the fuzzy set A is given by the linear combination*

$$A = \sum_{i=1}^n w_i A_i,$$

where $\sum_{i=1}^n w_i = 1$, then A is also sigmoid-shaped with the membership function $\sigma_a^{(\lambda)}(x)$, where

$$a = \sum_{i=1}^n w_i a_i,$$

$$\frac{1}{\lambda} = \sum_{i=1}^n w_i \frac{1}{\lambda_i}.$$

Proof: See Dombi (2009). ■

Theorem 1 allows us to separately aggregate the left hand sides and right hand sides of fuzzy intervals that are given by $\{a_l <_{(\lambda_l)} x <_{(\lambda_r)} a_r\}$ like inequalities.

The parameters a and λ of the sigmoid function $\sigma_a^{(\lambda)}$ can be unambiguously given by determining two points of the function curve. The sigmoid function neither takes the

value 0, nor the value 1, these are its limits. In practical applications, it may be useful if the function is given by two points which have vertical coordinates close to 0 and 1. Let ε be a small positive value, for example $\varepsilon = 0.001$, and

$$y_0 = \varepsilon$$

$$y_1 = 1 - \varepsilon.$$

If we wish $\sigma_a^{(\lambda)}$ to take the values of y_0 and y_1 at x_0 and x_1 , respectively, the parameters a and λ need to be set as follows:

$$a = \frac{x_0 \ln\left(\frac{1 - y_1}{y_1}\right) - x_1 \ln\left(\frac{1 - y_0}{y_0}\right)}{\ln\left(\frac{1 - y_1}{y_1}\right) - \ln\left(\frac{1 - y_0}{y_0}\right)}$$

$$\lambda = -\frac{\ln\left(\frac{1 - y_0}{y_0}\right)}{x_0 - a},$$

where $x_0 \neq a$. Let l, m, r be three (crisp) numbers so that $l < m < r$. If we require $\sigma_{a_l}^{(\lambda_l)}(x)$ to take the values of ε and $1 - \varepsilon$ at l and m , respectively, and we also require $\sigma_{a_r}^{(\lambda_r)}(x)$ to take the values of $1 - \varepsilon$ and ε at m and r , respectively, then $a_l, \lambda_l, a_r, \lambda_r$ need to be set as follows:

$$a_l = \frac{l + m}{2}$$

$$\lambda_l = \frac{2}{m - l} \ln\left(\frac{1 - \varepsilon}{\varepsilon}\right)$$

$$a_r = \frac{r + m}{2}$$

$$\lambda_r = \frac{2}{m - r} \ln\left(\frac{1 - \varepsilon}{\varepsilon}\right).$$

Next, if we generate the Dombi intersection of the increasing sigmoid fuzzy membership function $\sigma_{a_l}^{(\lambda_l)}(x)$ and the decreasing sigmoid fuzzy membership function $\sigma_{a_r}^{(\lambda_r)}(x)$ with the above parameters, then we get the membership function

$$\begin{aligned} \mu(x; l, m, r) &= \frac{1}{1 + e^{-\frac{2}{m - l} \ln\left(\frac{1 - \varepsilon}{\varepsilon}\right)\left(x - \frac{l + m}{2}\right)} + e^{-\frac{2}{m - r} \ln\left(\frac{1 - \varepsilon}{\varepsilon}\right)\left(x - \frac{r + m}{2}\right)}} \\ &= \frac{1}{1 + ((1 - \varepsilon)/\varepsilon)^{-\frac{2}{m - l}\left(x - \frac{l + m}{2}\right)} + ((1 - \varepsilon)/\varepsilon)^{-\frac{2}{m - r}\left(x - \frac{r + m}{2}\right)}}. \end{aligned}$$

It can be proven that the function $\mu(x; l, m, r)$ has its maximum at

$$c = \frac{(m - l)(r - m) \ln\left(\frac{r - m}{m - l}\right)}{2(r - l) \ln\left(\frac{1 - \varepsilon}{\varepsilon}\right)} + m.$$

Furthermore, $\mu(l; l, m, r) = \mu(r; l, m, r) = \varepsilon$; that is, the function $\mu(x; l, m, r)$ may be viewed as the membership function of the fuzzy number ‘approximately c ’. Moreover,

$$\lim_{\varepsilon \rightarrow 0} \frac{(m - l)(r - m) \ln\left(\frac{r - m}{m - l}\right)}{2(r - l) \ln\left(\frac{1 - \varepsilon}{\varepsilon}\right)} = 0,$$

that is, if ε is close to zero, then $c \approx m$. Since the parameter ε is a small positive number (e.g. $\varepsilon = 0.001$), the membership function $\mu(x; l, m, r)$ approximates quite well the fuzzy number ‘approximately m ’. Figure 3 shows an example of the fuzzy number ‘approximately m ’ given by the membership function $\mu(x; l, m, r)$.

3.2. A possibilistic representation of the perceived performance

We have shown that the function $\mu(x; l, m, r)$, for any x , approximates the truth of the $\{x = m\}$ soft equality quite well, so we can represent $\{x = m\}$ by the function $\mu(x; l, m, r)$. Since $\mu(x; l, m, r) \in (0, 1)$ for any x , and $\mu(x; l, m, r)$ is strictly monotonously increasing, if $x < c$; and it is strictly monotonously decreasing, if $x > c$, there are exactly two places, $x_l^{(\alpha)}$, and $x_r^{(\alpha)}$ ($x_l^{(\alpha)} < x_r^{(\alpha)}$), at which the membership function $\mu(x; l, m, r)$ takes the value of α , where $\alpha \in (0, 1)$. Thus, the α cut of $\mu(x; l, m, r)$ is the interval $(x_l^{(\alpha)}, x_r^{(\alpha)})$; that

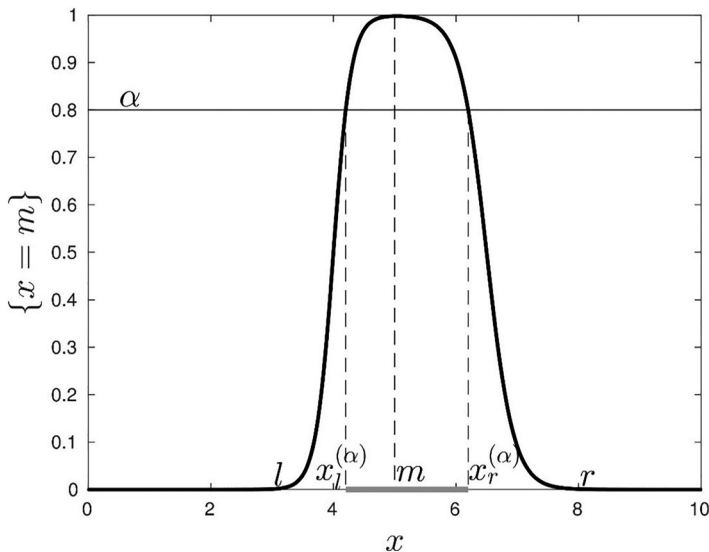


Figure 3. The α cut of the fuzzy number ‘approximately m ’.

is,

$$(x_l^{(\alpha)}, x_r^{(\alpha)}) = \{x: x \in R, \mu(x; l, m, r) \geq \alpha\}.$$

Owing to the construction of the function $\mu(x; l, m, r)$, the approximate values of $x_l^{(\alpha)}$, and $x_r^{(\alpha)}$ can be computed from the left-hand side and right-hand side sigmoid components of $\mu(x; l, m, r)$:

$$x_l^{(\alpha)} \approx -\frac{1}{\frac{2}{m-l} \ln\left(\frac{1-\varepsilon}{\varepsilon}\right)} \ln\left(\frac{1-\alpha}{\alpha}\right) + \frac{l+m}{2}$$

$$x_r^{(\alpha)} \approx -\frac{1}{\frac{2}{m-r} \ln\left(\frac{1-\varepsilon}{\varepsilon}\right)} \ln\left(\frac{1-\alpha}{\alpha}\right) + \frac{r+m}{2}.$$

When ‘approximately m ’ is utilised for describing a perceived performance, and the fuzzy number ‘approximately m ’ is given by the membership function $\mu(x; l, m, r)$, then the left- and right ends of the α cut interval of $\mu(x; l, m, r)$, respectively, $x_l^{(\alpha)}$, and $x_r^{(\alpha)}$, can be interpreted as follows. If the perceived performance is between $x_l^{(\alpha)}$ and $x_r^{(\alpha)}$, then the truth level of the statement that the performance is ‘approximately m ’ is at least α .

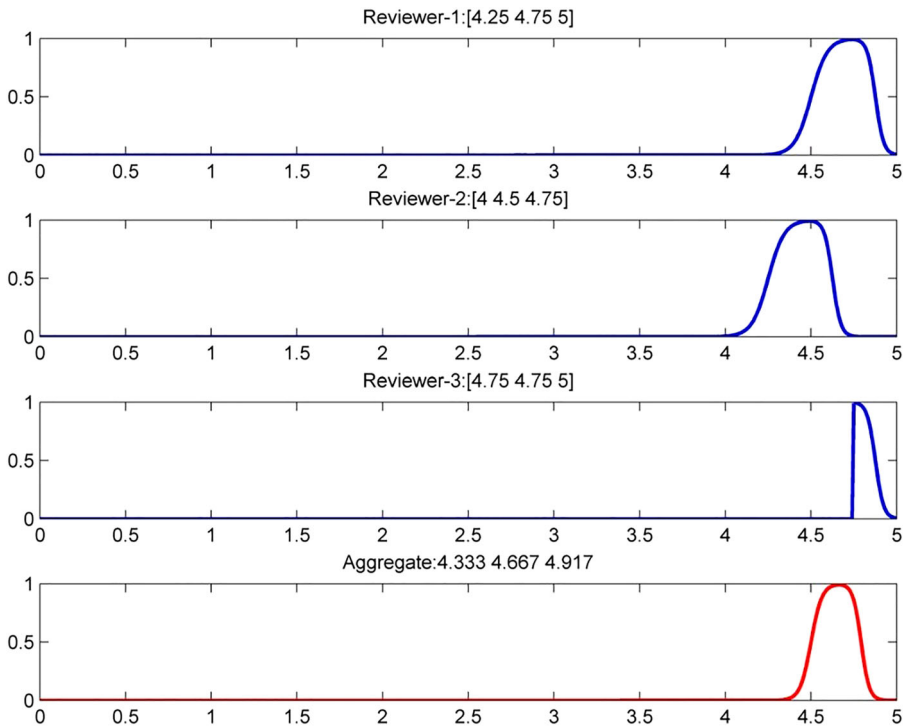


Figure 4. Fuzzy numbers based evaluation of dimension *Learning, explaining the terminology* in case of Lecturer 1.

Table 1. Parameters of fuzzy numbers, means and 95% confidence intervals of crisp evaluations representing the average evaluations in each dimension used to evaluate lecturers' performance.

Evaluation dimension	Fuzzy number			Crisp evaluation	
	Left	Centre	Right	Mean	St. deviation
Volume, intonation	4.127	4.441	4.642	4.568	0.545
Grammar, intelligibility, speech rate	4.268	4.583	4.771	4.636	0.532
Learning, explaining the terminology	4.004	4.646	4.717	4.682	0.471
Explanatory capability	3.997	4.244	4.657	4.295	0.553
Maintaining students' attention	3.797	4.418	4.587	4.432	0.625
Choosing the proper presentation techniques	3.557	4.198	4.447	4.273	0.949
Quality of the slide show	3.716	3.974	4.356	4.119	0.633
Consonance of the slide show with verbal communication	4.667	4.833	4.917	4.762	0.431
Logical structure of the lecture and presentation	4.663	4.917	4.988	4.818	0.446
Average evaluation	4.088	4.473	4.676	4.510	0.632

Figure 3 shows a graphical example of the above-mentioned interpretation of $x_l^{(\alpha)}$, and $x_r^{(\alpha)}$.

4. Case study

4.1. Fuzzy Likert scale based evaluation of teaching performance

Based on the favourable properties of fuzzy Likert scales, that can help to overcome the difficulties associated with the peer review process, some aspects of the applied

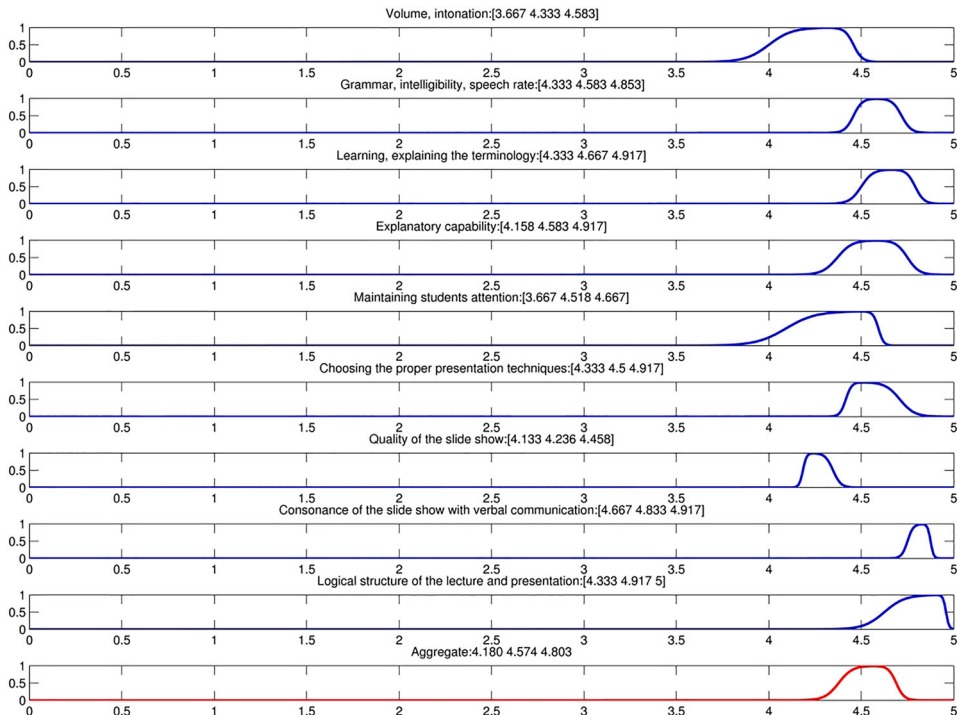


Figure 5. Fuzzy numbers based evaluations of Lecturer 1.

questionnaire have been evaluated by using a fuzzy scale parallel to the traditional Likert scale during the fall semester 2016. Based on various feedbacks, the evaluation dimensions related to the delivery of the lectures have been selected for this purpose, since these dimensions have turned out to be the most difficult to evaluate on a traditional scale. In case of five evaluated lecturers, three reviewers have been asked to express their judgement on a fuzzy Likert scale regarding the dimensions mentioned previously. The scale has been applied with a division of 0.25 units in order to allow the reviewers a more precise reflection of their judgement. Altogether 135 fuzzy evaluations have been collected representing three reviewers' opinion of performance of five selected reviewees in nine evaluation dimensions listed above. It can be concluded that the reviewers have not experienced any difficulties when making their evaluation; Gil, Lubiano, De Sáa, and Sinova (2015) have come to the same conclusion.

The following figure (Figure 4) illustrates the fuzzy numbers (represented by blue curves) used to evaluate Lecturer 1's performance as well as the aggregate fuzzy evaluation (denoted by the red curve) in the dimension entitled *Learning, explaining the terminology*.

The aggregate evaluations of the other dimensions in case of Lecturer 1 have been computed similarly. Figure 5 represents the aggregated fuzzy evaluations in each dimension as well as the average evaluation of the dimensions, which represents the average evaluation of Lecturer 1's performance.

The average evaluations for all involved lecturers are presented in Figure 6. Based on the evaluations shown in Figure 6, Lecturer 1, 2 and 5 has approximately the same performance, however, in case of Lecturer 1 and 2, the performance or the judgement of the reviewers is more unbalanced. Based on the traditional Likert scale-based evaluation,

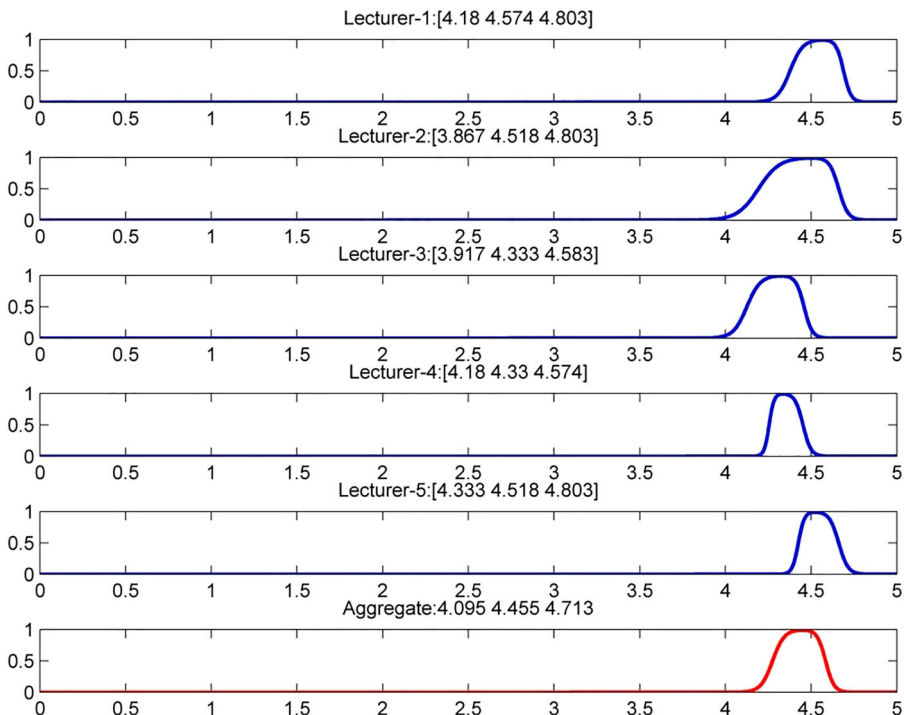


Figure 6. Aggregate evaluations of the 5 selected lecturers by fuzzy numbers.

however, it was almost impossible to make a distinction between their performance levels.

Aggregate evaluations can be computed not only for a given lecturer, but for a specific dimension as well. Based on the evaluation of the five involved lecturers, the average judgements in each dimension are presented and listed in Figure 7. Table 1 contains the parameters of fuzzy numbers and the average of crisp evaluations given on a traditional Likert scale by the same reviewers in each dimension. It can be concluded that the means of crisp evaluations and the centres of fuzzy numbers do not differ significantly.

4.2. Discussion of findings

The main benefit of applying fuzzy numbers to evaluate lecturers' performance is their ability to model the imprecision and the uncertainty inherent in reviewers' evaluation as well the representation of the variation of performance. Based on Figures 4–7 it can be concluded that the performances of the observed lecturers are unbalanced and reviewers have different perceptions on their performance as well. Utilising traditional Likert scale based evaluation, it is almost impossible to capture accurately this variation of the performance (see Table 1 and Figure 6). As the case study demonstrates, the centres of fuzzy numbers do not differ significantly from the means of crisp evaluations. Despite the fact that fuzzy evaluations do not differ more than half a unit, the information encoded in fuzzy evaluations supports well the identification of areas need to be improved. Moreover, the evaluation dimensions, which turned out to be more unbalanced, were also downgraded by students, further emphasising the need for improvement in these areas. Having

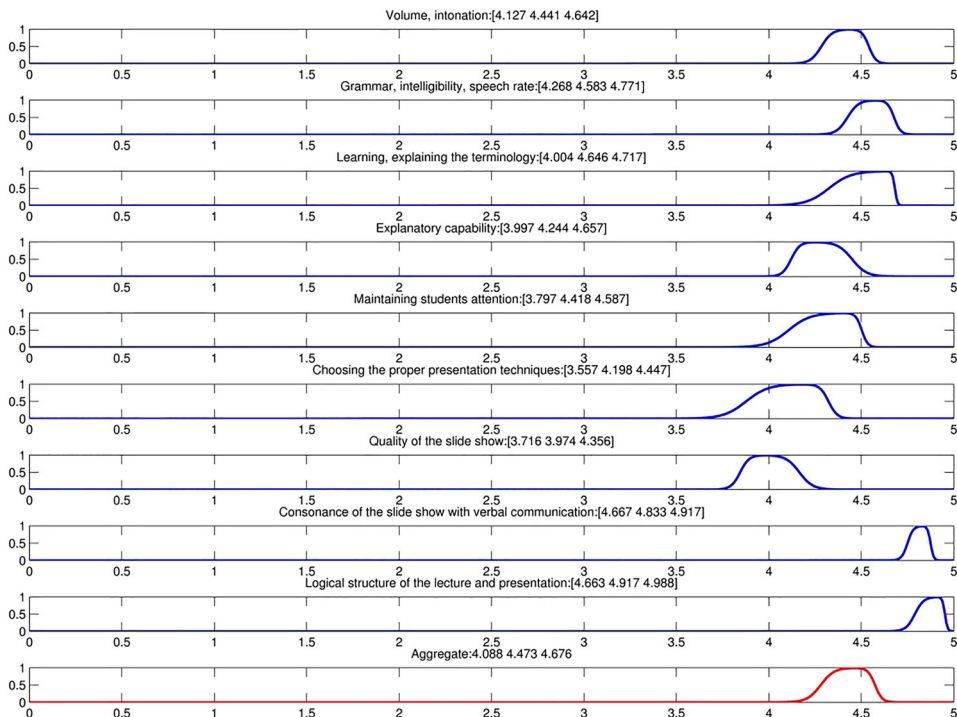


Figure 7. The average evaluations of five involved lecturers' performance in each dimension.

investigated the evaluation dimensions in case of which fuzzy numbers indicate a greater diversity and taking the narrative feedbacks of previous evaluations into account, it can be concluded that the variation of the observed lecturer's performance could be expressed on a fuzzy scale, since the fuzzy numbers are spread out in those dimensions in which the previous narrative comments indicated a more unbalanced performance. That is, by offering the opportunity to use a fuzzy Likert scale to evaluate the observed lecture, reviewers can express their uncertainty and the variability of the observed colleague's performance during a single lecture in a quantitative way.

5. Conclusions and managerial implications

Based on the first results it can be concluded that the presented fuzzy judgement can help the university staff to overcome several difficulties associated with peer evaluations. It is beneficial from the faculty's point of view, namely, it allows comparisons between different lecturers' performance and the identification of best practices and weaknesses as well. Based on the encouraging results, applying fuzzy evaluation in the peer review process may be considered as a viable technique.

5.1. Practical implications

By applying fuzzy numbers for peer reviewing, the faculty can decrease the amount of narrative comments as well, which is beneficial from the statistical point of view of the analyses and allows more information to be taken into account when comparing different lecturers' performance and identifying best practices.

Besides peer evaluation, the application of the presented methodology may arise in evaluations where Likert-type scales are applied traditionally. The most desirable way is to extend the application of fuzzy scales to student satisfaction measurement, since students face similar difficulties when evaluating courses, which results not only in a low response rate, but also in less reliable information. Moreover, students usually have various perceptions on teaching quality which cannot be measured by a single number on a traditional scale. By applying fuzzy numbers, students can choose more than one single value which represents their opinion, for example, their optimistic and pessimistic judgement on the teachers' performance.

5.2. Limitations and further research plans

Besides feeding back the results to the peer reviewed lecturers, student satisfaction or course evaluation results should be continually monitored and compared to the results of the peer review programme. In order to evaluate the ultimate impact of peer reviewing on teaching performance and the effectiveness of fuzzy Likert scales used for this purpose, relationships between the results of peer reviewing and student satisfaction surveys should be further investigated. Extending fuzzy evaluations to student satisfaction may result in a more detailed investigation of cause and effect relationships as well. Fuzzy regression (Alfonso, López, de Hierro, & Roldán, 2016) or fuzzy inference systems (e.g. Hammed, 2011; Jónás & Árrva, 2016) would be a suitable tool to deal with this challenge.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Alfonso, G., López, R., de Hierro, A. F., & Roldán, C. (2016). A fuzzy regression model based on finite fuzzy numbers and its application to real-world financial data. *Journal of Computational and Applied Mathematics*, 24(2), 344–359.
- Amini, S., & Jochem, R. (2011). *A conceptual model based on the fuzzy set theory to measure and evaluate the performance of service processes*. Paper presented at enterprise distributed object computing conference workshops (EDOCW), 29 Aug–2 Sept 2011, Helsinki, Finland, pp. 122–131.
- Andayani, S., Hartati, S., Wardoyo, R., & Mardapi, D. (2017). Decision-making model for student assessment by unifying numerical and linguistic data. *International Journal of Electrical and Computer Engineering*, 7(1), 363–373.
- Başaran, M. A., Kalaycı, N., & Atay, M. T. (2011). A novel hybrid method for better evaluation: Evaluating university instructors teaching performance by combining conventional content analysis with fuzzy rule based systems. *Expert Systems with Applications*, 38(10), 12565–12568.
- Battisti, F., Nicolini, G., & Salini, S. (2005). The Rasch model to measure service quality. *The ICFAI Journal of Services Marketing*, 3(3), 58–80.
- Battisti, F., Nicolini, G., & Salini, S. (2010). The Rasch model in customer satisfaction survey data. *Quality Technology & Quantitative Management*, 7(1), 15–34.
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48–62.
- Blackmore, J. A. (2005). A critical evaluation of peer review via teaching observation within higher education. *International Journal of Educational Management*, 19(3), 218–232.
- Brent, R., & Felder, R. M. (2004). A protocol for peer review of teaching. *Education designs*, North Carolina State University, Session 3530.
- Brochado, A. (2009). Comparing alternative instruments to measure service quality in higher education. *Quality Assurance in Education*, 17(2), 174–190.
- Büyükközkcan, G., Ruan, D., & Feyzioğlu, O. (2007). Evaluating e-learning web site quality in a fuzzy environment. *International Journal of Intelligent Systems*, 22(5), 567–586.
- Cabrerizo, F. J., López-Gijón, J., Martínez, M. A., Morente-Molinera, J. A., & Herrera-Viedma, E. (2017). A fuzzy linguistic extended LibQUAL+ model to assess service quality in academic libraries. *International Journal of Information Technology & Decision Making*, 16(1), 225–244.
- Calcagni, A., & Lombardi, L. (2014). Dynamic fuzzy rating tracker (DYFRAT): A novel methodology for modeling real-time dynamic cognitive processes in rating scales. *Applied Soft Computing*, 24, 948–961.
- Carrasco, R. A., Villar, P., Hornos, M. J., & Herrera-Viedma, E. (2011). A linguistic multi-criteria decision making model applied to the integration of education questionnaires. *International Journal of Computational Intelligence Systems*, 4(5), 946–959.
- Chang, T. C., & Wang, H. (2016). A multi criteria group decision-making model for teacher evaluation in higher education based on cloud model and decision tree. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(5), 1243–1262.
- Chen, T. C. (2001). Applying linguistic decision-making method to deal with service evaluation problems. *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems*, 9(1), 103–114.
- Courmeya, C. A., Pratt, D. D., & Collins, J. (2008). Through what perspective do we judge the teaching of peers? *Teaching and Teacher Education*, 24, 69–79.
- Deng, W. J. (2008). Fuzzy importance-performance analysis for determining critical service attributes. *International Journal of Service Industry Management*, 19(2), 252–270.
- de Sáa, S. D. L. R., Gil, M.Á., 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(1), 111–126.
- De Witte, K., & Rogge, N. (2011). Accounting for exogenous influences in performance evaluations of teachers. *Economics of Education Review*, 30(4), 641–653.
- Dombi, J. (2008). Towards a general class of operators for fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 16(2), 477–484.
- Dombi, J. (2009). Pliant arithmetics and pliant arithmetic operations. *Acta Polytech Hun*, 6(5), 19–49.

- Ellis, L., Burke, D. M., Lomire, P., & McCormack, D. R. (2003). Student grades and average ratings of instructional quality: The need for adjustment. *The Journal of Educational Research*, 97(1), 35–40.
- Frühwirth-Schnatter, S. (1992). On statistical inference for fuzzy data with applications to descriptive statistics. *Fuzzy Sets and Systems*, 50, 143–165.
- Gil, M. Á., & González-Rodríguez, G. (2012). Fuzzy vs. Likert scale in statistics. In E. Trillas, P. P. Bonissone, L. Magdalena, & J. Kacprzyk (Eds.), *Combining experimentation and theory* (pp. 407–420). Berlin, Heidelberg: Springer.
- Gil, M. Á., Lubiano, M. A., De Sáa, S. D. L. R., & Sinova, B. (2015). Analyzing data from a fuzzy rating scale-based questionnaire. A case study. *Psicothema*, 27(2), 182–191.
- Hammed, I. A. (2011). Using Gaussian membership functions for improving the reliability and robustness of students' evaluation systems. *Expert Systems with Applications*, 38, 7135–7142.
- Hartley, J. (2014). Some thoughts on Likert-type scales. *International Journal of Clinical and Health Psychology*, 14(1), 83–86.
- Herrera, F., & Herrera-Viedma, E. (2000). Choice functions and mechanisms for linguistic preference relations. *European Journal of Operational Research*, 120(1), 144–161.
- Herrera, F., López, E., Mendana, C., & Rodríguez, M. A. (1999). Solving an assignment–selection problem with verbal information and using genetic algorithms. *European Journal of Operational Research*, 119(2), 326–337.
- Hesketh, B., Pryor, R., Gleitzman, M., & Hesketh, T. (1988). Practical applications and psychometric evaluation of a computerized fuzzy graphic rating scale. *Advances in Psychology*, 56, 425–454.
- Ihsan, A. K. A. M., Taib, K. A., Talib, M. Z. M., Abdullah, S., Husain, H., Wahab, D. A., ... Abdul, N. A. (2012). Measurement of course evaluation for lecturers at the faculty of engineering and built environment. *Procedia – Social and Behavioral Sciences*, 60, 358–364.
- Jónás, T., & Árva, G. (2016). Application of fuzzy inference systems build from data for quality and service management purposes. In S. M. Dahlgaard-Park & J. J. Dahlgaard (Eds.), *19th QMOD-ICQSS conference international conference on quality and service sciences*. Roma, Italy, 21.09.2016–23.09.2016. Lund University Library Press, Lund, pp. 519–534.
- Kacprzyk, J. (1986). Towards a 'human-consistent' multistage decision making and control models using fuzzy sets and fuzzy logic. *Fuzzy Sets and Systems*, 18(3), 299–314.
- Kuzmanovic, M., Savic, G., Popovic, M., & Martic, M. (2013). A new approach to evaluation of university teaching considering heterogeneity of students' preferences. *Higher Education*, 66(2), 153–171.
- Lalla, M., Facchinetti, G., & Mastroleo, G. (2005). Ordinal scales and fuzzy set systems to measure agreement: An application to the evaluation of teaching activity. *Quality & Quantity*, 38(5), 577–601.
- Li, Q. (2013). A novel Likert scale based on fuzzy sets theory. *Expert Systems with Applications*, 40(5), 1609–1618.
- Liaw, S. H., & Goh, K. L. (2003). Evidence and control of biases in student evaluations of teaching. *International Journal of Educational Management*, 17(1), 37–43.
- Lin, H. F. (2010a). An application of fuzzy AHP for evaluating course website quality. *Computers & Education*, 54(4), 877–888.
- Lin, H. T. (2010b). Fuzzy application in service quality analysis: An empirical study. *Expert Systems with Applications*, 37(1), 517–526.
- Liou, T. S., & Chen, C. W. (2006). Subjective appraisal of service quality using fuzzy linguistic assessment. *International Journal of Quality & Reliability Management*, 23(8), 928–943.
- Liu, R., Cui, L., Zeng, G., Wu, H., Wang, C., Yan, S., & Yan, B. (2015). Applying the fuzzy SERVQUAL method to measure the service quality in certification and inspection industry. *Applied Soft Computing*, 26, 508–512.
- Lozano, L. M., García-Cueto, E., & Muñoz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73–79.
- Lubiano, M. A., de Sáa, S. D. L. R., Montenegro, M., Sinova, B., & Gil, M. Á. (2016). Descriptive analysis of responses to items in questionnaires. Why not using a fuzzy rating scale? *Information Sciences*, 360, 131–148.
- Lupo, T. (2013). A fuzzy ServQual based method for reliable measurements of education quality in Italian higher education area. *Expert Systems with Applications*, 40(17), 7096–7110.
- Lupo, T. (2016). A fuzzy framework to evaluate service quality in the healthcare industry: An empirical case of public hospital service evaluation in Sicily. *Applied Soft Computing*, 40, 468–478.

- Mashhadiabdol, M., Sajadi, S. M., & Talebi, K. (2014). Analysis of the gap between customers' perceptions and employees' expectations of service quality based on fuzzy SERVQUAL logic (case study: Mofid children's hospital in Tehran, Iran). *International Journal of Services and Operations Management*, 17(2), 119–141.
- Murray, J. (2013). Likert data: What to use, parametric or non-parametric. *International Journal of Business and Social Science*, 4(11), 258–264.
- Nadiri, H., Kandampully, J., & Hussain, K. (2009). Students' perceptions of service quality in higher education. *Total Quality Management*, 20(5), 523–535.
- Quirós, P., Alonso, J. M., & Pancho, D. P. (2016). Descriptive and comparative analysis of human perceptions expressed through fuzzy rating scale-based questionnaires. *International Journal of Computational Intelligence Systems*, 9(3), 450–467.
- Rouyendegh, B. D., & Erkan, T. E. (2013). An application of the fuzzy ELECTRE method for academic staff selection. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 23(2), 107–115.
- Samson, S., & McCrea, D. E. (2008). Using peer review to foster good teaching. *Reference Services Review*, 36(1), 61–70.
- Sowa, J. F. (2013). What is the source of fuzziness? In R. Seiging, E. Trillas, C. Moraga, & S. Termini (Eds.), *On fuzziness* (pp. 645–652). Berlin, Heidelberg: Springer.
- Teeroovengadam, V., Kamalanabhan, T. J., & Seebaluck, A. K. (2016). Measuring service quality in higher education: Development of a hierarchical model (HESQUAL). *Quality Assurance in Education*, 24(2), 244–258.
- Tóth Zs, E., Andor, G., & Árva, G. (2017a). Peer review of teaching at Budapest University of technology and economics - faculty of economic and social sciences. *International Journal of Quality and Service Sciences*, 9(3/4), 402–424.
- Tóth Zs, E., Surman, V., & Árva, G. (2017b). Challenges in course evaluations at Budapest University of Technology and Economics. In Z. Bekirogullari, M. Y. Minas, & R. X. Thambusamy (Eds.), *8th ICEEPSY - International conference on education and educational psychology*. Porto, Portugal, 2017.10.11–2017.10.14. Future Academy, 2017, pp. 629–641.
- Washer, P. (2006). Designing a system for observation of teaching. *Quality Assurance in Education*, 14(3), 243–250.
- Yu, C. M., Tsang, H. T., & Chen, K. S. (2016). Developing a performance evaluation matrix to enhance the learner satisfaction of an e-learning system. *Total Quality Management & Business Excellence*, 1–19. Published online 19 Sept 2016. <http://doi.org/10.1080/14783363.2016.1233809>.
- Zhang, J., Lin, T., & Ren, L. (2010). *Dynamic fuzzy evaluation for e-commerce service quality based on the SERVPERF*. Paper presented at the International Conference on E-Business and E-Government 2010 (ICEE), pp. 576–579. Retrieved from <http://ieeexplore.ieee.org/document/5590689/>