

8th ICEEPSY 2017
**The International Conference on Education and Educational
Psychology**

**ENHANCING THE RELIABILITY OF MEASUREMENTS AND
EVALUATIONS BASED ON SERVICE QUALITY MODELS**

Zsuzsanna Eszter Tóth (a)*, Gábor Árva (b), Vivien Surman (c)

*Corresponding author

- (a) Department of Management and Corporate Economics, Budapest University of Technology and Economics,
Magyar tudósok körútja 2, H-1117 Budapest, Hungary, toth.zsuzsanna.eszter@gmail.com
(b) Department of Management and Corporate Economics, Budapest University of Technology and Economics,
Magyar tudósok körútja 2, H-1117 Budapest, Hungary, arva@mvt.bme.hu
(c) Department of Management and Corporate Economics, Budapest University of Technology and Economics,
Magyar tudósok körútja 2, H-1117 Budapest, Hungary, surman@mvt.bme.hu

Abstract

Problem statement: The application of different service quality measurements and evaluations in higher educational context and the utilization and reliability of results are examined.

Research questions: What are the difficulties in the measurement and evaluation of specific service quality dimensions? How can the reliability of such service quality measurements be enhanced?

Purpose of the study: This paper introduces the application and the results of a student satisfaction questionnaire based on fuzzy Likert scale used for evaluating lecturers' performance at the Budapest University of Technology and Economics.

Research methods: In order to capture and accurately measure the diversity, subjectivity and imprecision inherent to students' evaluations, a methodology based on fuzzy numbers having sigmoid membership functions has been proposed. By applying the principles of Dombi's Pliant Arithmetics, the evaluations can be aggregated and statistically analysed in a convenient way.

Findings: Fuzzy-numbers based questionnaires result in a more precise reflection of human thinking and judgement. The Pliant Arithmetic-based approach allows us to aggregate the parameters of the left and right hand sides of the fuzzy number separately. This property results in a much simpler statistical analysis of the gathered data than the methods previously proposed in the literature.

Conclusions: The proposed methodology can be utilized in two ways. It can either be applied to identify lecturers' strengths and weaknesses in order to develop their teaching skills or to compare and analyse various students' evaluations. The results can serve as a base for establishing teaching regulations and discovering best practices as well.

© 2017 Published by Future Academy www.FutureAcademy.org.UK

Keywords: Service quality, Likert scale, fuzzy approach



1. Introduction

Nowadays, tertiary institutions are being called to account for the quality of educational services that they provide. While more accountability in higher education is desirable, the tools and mechanisms for its achievement are part of a hot debate in the relevant literature. The following questions naturally arise when it comes to the assessment of higher educational services. How can higher education (HE) institutions assess the quality of educational services they offer? How can they know reliably whether the expectations of customers and stakeholders of HE, primarily of students are met or not? If HE institutions wish to answer the previously addressed questions, they need suitable monitoring procedures and reliable methodologies to evaluate service quality and to identify the appropriate measure units to evaluate the achieved service performance level (Lupo, 2013, De Battisti et al., 2005; 2010) as HE services cannot be controlled or measured by classical measuring techniques and conventional measure units.

Measuring service quality and the satisfaction of stakeholders including students in HE is mainly realized through the application of Likert scales. The widely applied service quality models in HE including SERVQUAL (De Oliveira et al., 2009; Yousapronpaiboon, 2014), SERVPERF (Bayraktaroglu and Atrek, 2010; Brochado, 2009) and HEdPERF (Abdullah, 2005, 2006a, 2006b) are all based on 7-point Likert scales that vary from 1 (strongly disagree) to 7 (strongly agree) using different service quality domains. Measuring satisfaction in case of most HE stakeholders and peer evaluations are also based on the utilization of the traditional Likert scale (Gruber et al., 2010; Douglas and Douglas, 2006; Liu and Carless, 2006).

2. Problem Statement

Over the past decades there has been a trend toward fostering ratings with different data sources of teaching performance which could serve and broaden the evidence base used to evaluate courses and assess the quality of teaching. To serve this purpose, our Faculty has developed an internal quality enhancement system in order to further develop its teaching programs and practices in the academic year 2015/2016. Based on the relevant literature and taking international practices into consideration, a semester-long peer review of teaching program has been launched at the Faculty. The questionnaires applied to evaluate lecturers' classroom performance include the observation of lectures, midterm tests and/or exams. Since the students' satisfaction is highly dependent on the methods used to evaluate their performance during the semester, the peer review program has also been completed by the evaluation of midterm tests and exams by students right after the midterm test or exams (besides the aforementioned end of the term course evaluations). This kind of feedback is of high importance not only from the observed lecturer's, but from the faculty's point of view as well. The questionnaires filled out both by peer reviewees and students consist of two main parts: a numerical scale assessment in case of which most of the aspects of the lecturer's performance is evaluated using a traditional Likert-scale and there is a second part standing for narrative comments. Despite the advantages of the peer review program and the positive feedbacks coming from the participants, three main problems of student feedbacks have been found that can affect the reliability of results originating from the process of peer reviewing, namely, the

uncertainty inherent among the evaluations, the variation of lecturers' performance within the semester and the lack of methods available to compare different students' narrative evaluations.

3. Research Questions

Recent literature argues that the use of Likert scales reduces the subtlety of human perceptions as individuals can hardly use an exact number to express their opinion about a given situation. As an alternative, linguistic assessment is preferred to represent that specific numerical value (Herrera and Herrera-Viedma, 2000; Herrera et al., 1999; Kacprzyk, 1986; Chen, 2001). 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.

In designing questionnaires concerning variables which cannot be directly measured by means of exact numerical values but can be graded to some extent (like perceived service quality, satisfaction, perception, attitude), commonly employed scales are Likert ones. It is a discrete scale by choosing the most appropriate 'values' within a class according to the rater's judgement, opinion, valuation (Gil and González-Rodríguez, 2012) and lead to ordinal data from a set of pre-fixed categories. 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 et al., 2008).

Rating items in a questionnaire can be considered as a complex task (Jonessen, 2000) as raters make multiple decisions under uncertainty. Likert-type scales have several weaknesses and do not ease the task of the rater (Jamieson, 2004; Carifio and Perla, 2007). The number of 'values' to choose from is small (Gil and 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 et al., 2016). To some extent the ideal solution would be increasing the number of choices, but it cannot be achieved by using a natural language (Sowa, 2013). If the aim is to exploit individual differences in responding to questionnaires, there is a need for a rich and expressive scale in "something can be meaningful although we cannot name it" (Ghneim, 2013). To manage these disadvantages there is an alternate approach which takes into account that the nature of most attributes concerning evaluations, judgements involve subjectivity and certain imprecision. One of our research questions arose here: Is it possible from methodological point of view to handle the aforementioned problems in our peer evaluation program by alternate approaches?

Hesketh et al. (1988) proposed the fuzzy rating scale without respondents being constrained to choose among a few pre-specified questions. This kind of scale has the ability to model the imprecision of human rating evaluations, formalize them mathematically, to 'precisiate' them in a continuous way, and to develop mathematical computation with them (de Sáa et al., 2015; Calcagni and Lombardi, 2014; Gil and González-Rodríguez, 2012). This approach leads to a fuzzy-valued response format enabling a

variability and accuracy which would not be captured when using a Likert scale. The fuzzy-scale is rich and expressive enough to find a value in it fitting appropriately the valuation, opinion, judgement involving subjective perceptions in most real life situations. Fuzzy rating scales have been intensively applied in higher education context to measure quality related issues (see e.g. Basaran et al., 2011; Lalla et al., 2005; Yu et al., 2016; Lupo, 2013; Liu et al., 2015; Venkatesan and Fragomeni, 2008). Based on the relevant literature the next research issue was addressed by applying a fuzzy-scale in case of student evaluations in the framework of the peer review program and analyze the benefits of fuzzy scales compared to traditional Likert scales.

4. Purpose of the Study

Since students are the most important stakeholders of higher education and they are those who have direct interactions with the lecturers, they are considered to be the most reliable source of information regarding quality assessments. Some problems arise in connection with student satisfaction measurement. Usually, students take into account their relationship with the lecturer when judging teaching quality, even if the goal of this measurement is to gain objective information. The own opinion of students is strongly influenced by others and students tend to express a “common opinion” on quality. The third problem inherent in student evaluations is the contrasting perceptions as time goes on. Students have quite different feelings right after the midterm test, after getting to know their results and after successfully passing the course, that is, their judgement on teaching quality is continuously reconsidered. All of the three factors lead to different but parallel existing perceptions and opinions. Moreover, the performance of a lecturer is often unbalanced and fluctuates during the lecture or as the semester goes on. In these situations, it could be quite difficult to choose a single number which can depict the performance. If respondents are constrained to choose a given number on a traditional Likert scale, it could be assumed that this number will represent an average performance during the whole lecture or regarding all dimensions along which lecturers’ performance are evaluated. This average performance is seldom representative and not always sufficient enough to identify strengths and weaknesses. Moreover, the retrospective statistics including the mean, the range and the standard deviation computed based on these evaluations are more likely to reflect the differences between the various students’ judgements than the variability of the lecturer’s performance.

Narrative comments are of high importance and kindly welcomed from the lecturers’ points of view. They emphasize the evidence on which students base their evaluations and the selection of lecturers’ strengths and weaknesses. They can reflect the variability of the observed lecturer’s performance or the contrasting perceptions of students as well. On the contrary, these narrative comments are difficult to analyse and the lack of simple methods to deal with linguistic feedbacks led to an insufficient elaboration of these kinds of feedbacks. It means that by comparing and evaluating different lecturers’ performance based only on the 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.

This paper focuses on a challenging problem which is related to how to handle properly the inherent uncertainty of human perceptions. Namely, we illustrate new ways to interpret and analyse fuzzy

data coming out from a special case of survey, the so-called fuzzy rating scale-based questionnaire applied in the peer review program at our Faculty to evaluate lecturers' teaching performance. The proposed fuzzy Likert scale can help to overcome the aforementioned difficulties. This approach can help to deal with vagueness arising either from uncertainty of the students or from the fluctuation of the observed lecturer's performance. By providing a fuzzy Likert scale to evaluate the lecture, students can express their uncertainty, their contrasting perceptions and the variability of the observed lecturer's performance in a quantitative way. 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. Following Dombi's Pliant Inequality Model (Dombi, 2009) and Theorem 1, introduced in Section 3.1, 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.

5. Research Methods

In the followings the theoretical background of the proposed methodology is shortly discussed.

5.1. 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.

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

- 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 1 shows examples of sigmoid function graphs. Let us assume, that we have two sigmoid functions, $\sigma_{a_l}^{(\lambda_l)}(x)$ and $\sigma_{a_r}^{(\lambda_r)}(x)$, where $\lambda_l > 0$ and $\lambda_r < 0$, that is, $\sigma_{a_l}^{(\lambda_l)}(x)$ is strictly monotonously increasing, while $\sigma_{a_r}^{(\lambda_r)}(x)$ is strictly monotonously decreasing. Then, we will use the Dombi conjunction operator (Dombi intersection) to implement intersection of two fuzzy sets (Dombi, 2008).

Definition 2. 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 $\mu_{A_1 \cap A_2}(x)$:

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

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

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

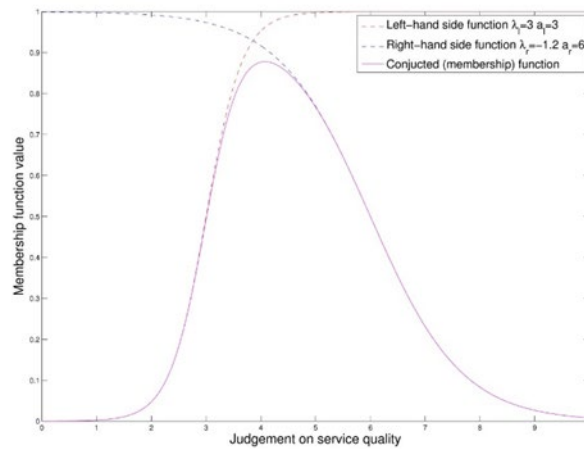


Figure 01. Intersection of two fuzzy sets given by an increasing and a decreasing sigmoid membership function

Utilizing Definition 1 we get:

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

Figure 01. shows the intersection of two fuzzy sets given by an increasing and a decreasing sigmoid membership function. The following theorem allows us to separately aggregate the left hand sides and right hand sides of fuzzy numbers.

Theorem 1. If A_1, A_2, \dots, A_n are fuzzy sets with the membership functions $\sigma_{a_1}^{(\lambda_1)}, \sigma_{a_2}^{(\lambda_2)}, \dots, \sigma_{a_n}^{(\lambda_n)}$, respectively, $sgn(\lambda_1) = sgn(\lambda_2) = \dots = sgn(\lambda_n)$, 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 $\sigma_a^{(\lambda)}$, 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).

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:

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

where $x_0 \neq a$.

This approach enables us to represent the “approximately m ” value by the parameter triple l, m, r . Namely, “approximately m ” can be given as the Dombi intersection of the increasing sigmoid fuzzy membership function $\sigma_{a_l}^{(\lambda_l)}(x)$ and the decreasing sigmoid fuzzy membership function $\sigma_{a_r}^{(\lambda_r)}(x)$, where

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

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

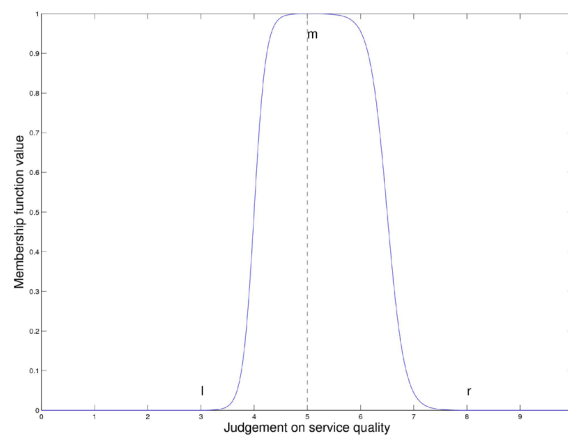


Figure 02. The "approximately m " given as intersection of two sigmoid membership functions

Figure 02. shows an example of the fuzzy number “approximately m ” that is composed of a left hand side and a right hand side sigmoid functions. Based on the favourable properties of fuzzy Likert-scales that can help to overcome the difficulties associated with student satisfaction measurement, a pilot fuzzy number based evaluation has been launched at the Faculty during the fall semester 2016. In order to gain experiences with fuzzy evaluation of teaching performance, the questionnaires used to evaluate the midterm tests/exams by students have been selected to be evaluated by a fuzzy Likert scale. In case of 5 courses 7 midterm tests have been selected after which students have been asked to use a fuzzy Likert scale to evaluate the midterm test. The eight dimensions used to evaluate the midterm tests are as follows: *D1 availability of instructional materials, D2 midterm test, exam circumstances, D3 review the course of tests, exams, D4 clarity of exam questions, D5 consonance of exam questions with requirements, D6 clarity of result calculation, D7 standard of consultation opportunities, D8 standard of midterm test/exam viewing opportunities.* The scale was applied with a division of 0.25 units in order to allow students a more detailed reflection of their judgement. After each midterm test, 10-15 students evaluated the lecturers’ performance on this fuzzy scale, so that altogether 85 fuzzy questionnaires have been filled out, while most of the students did not experience any difficulties when making their evaluation. The experiences were in accordance with the results of Gil et al. (2015). Parallel to the fuzzy number based evaluation, the traditional Likert scale based questionnaires have been launched as well, which allowed us to compare the results.]

6. Findings

Figure 03. shows the fuzzy evaluation of Lecturer 1; the fuzzy numbers in blue depict a student's judgement in each dimension while the red number denotes the average performance of the given lecturer which was computed according to Dombi's model, introduced in section 5.1, based on the performance in each dimension.

In case of Lecturer 1 altogether 12 students have been asked to use the fuzzy Likert scale to evaluate the lecturer's performance. Figure 04. shows the average evaluations of these 12 students in each dimension (denoted by fuzzy numbers in blue) as well as the average performance of this lecturer based on the 12 students' evaluations (depicted by the fuzzy numbers in red).

Table 01. contains the parameters of the fuzzy numbers representing the average evaluation in each dimension and the boundaries of the 95% confidence interval of the expected values of the crisp evaluations as well.

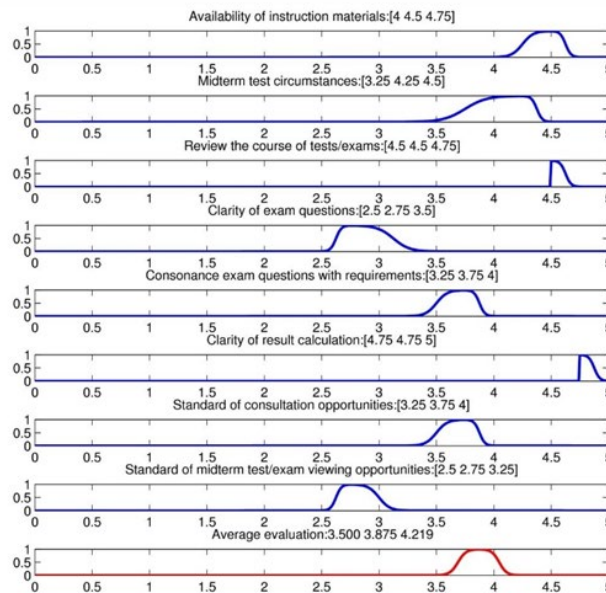


Figure 03. A student's evaluation in each dimension used to judge the midterm test as well as the average performance of Lecturer 1

It can be concluded that the centre of the fuzzy numbers representing the most probable response and the mean of the crisp evaluations do not differ significantly. The width and the location of the 95% confidence interval of the expected value of the crisp evaluation usually coincides with the fuzzy evaluation; however, in case of average crisp evaluations, that are estimated based on a larger sample, the confidence interval becomes tighter. It should be emphasized, however, that there is no mathematical connection between fuzzy and crisp evaluations; while fuzzy evaluations represent a possibilistic, crisp judgements appear for a probabilistic approach.

Similar to Lecturer 1, the average performance in each dimension have been computed for the other 4 lecturers as well. Based on those evaluations, the average performance of the lecturers is shown in Figure 05. Figure 05. demonstrates well the benefits of fuzzy number based evaluation. Lecturer 4 and 5

have almost the same performance (the centre of the fuzzy numbers equals 4.141 and 4.133, respectively), however, either the performance or the judgement is more unbalanced in case of Lecturer 5. On the contrary, based on the traditional Likert scale based evaluation it is almost impossible to make a difference between their performance, since Lecturer 4's average performance is 4.026 with the standard deviation of 0.897, while in case of Lecturer 5 the mean and the standard deviation of the crisp evaluations equals to 4.017 and 0.921, respectively. That is, fuzzy numbers contain more information than the traditional crisp approach. The reason for this is the fact that fuzzy number based questionnaires can depict the opinion of students more precisely, since this approach is able to depict the variation of the lecturer's performance during the semester or the various perceptions of students.

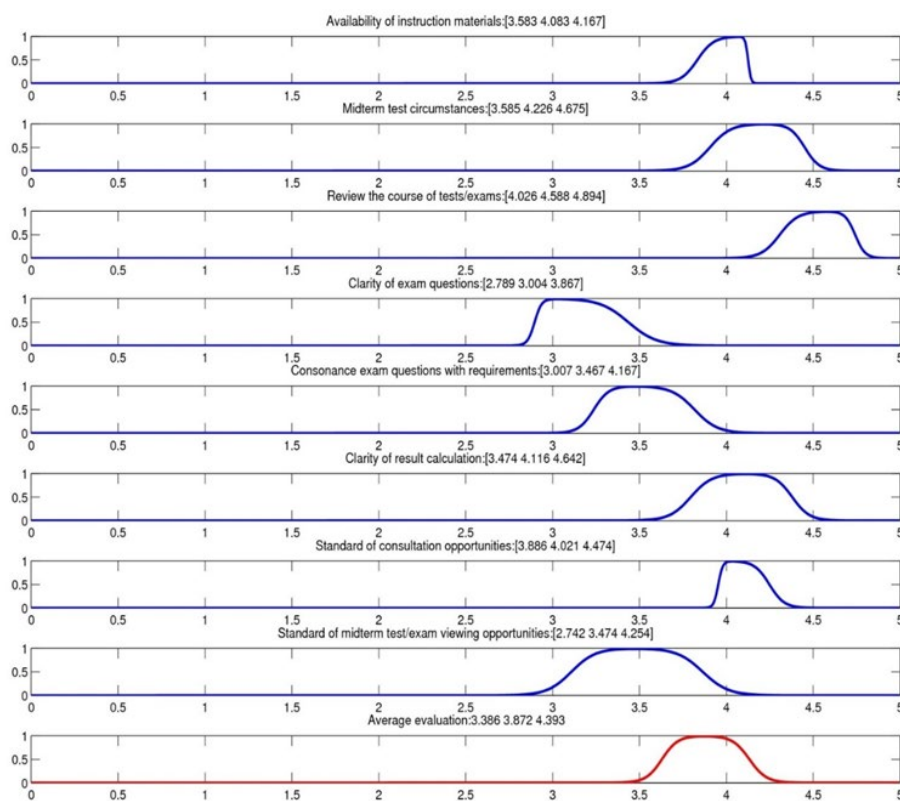


Figure 04. Average performance of Lecturer 1 in each dimension

Table 01. Parameters of the Fuzzy numbers and the boundaries of the 95% confidence interval used to evaluate Lecturer 1's performance in each dimension

Evaluation dimension	Fuzzy number			Crisp evaluation			
	Left	Centre	Right	Mean	St. deviation	Lower 95 %	Upper 95 %
Availability of instructional materials	3.583	4.083	4.167	3.917	0.515	3.589	4.244
Midterm test, exam circumstances	3.585	4.226	4.675	4.167	0.835	3.636	4.697
Review the course of tests, exams	4.026	4.588	4.894	4.583	0.669	4.159	5.008
Clarity of exam questions	2.789	3.004	3.867	3.250	1.055	2.579	3.921
Consonance of exam	3.007	3.467	4.167	3.583	0.996	2.950	4.216

questions with requirements							
Clarity of result calculation	3.474	4.116	4.642	4.083	0.996	3.450	4.716
Standard of consultation opportunities	3.886	4.021	4.474	3.917	1.165	3.177	4.657
Standard of midterm test/exam viewing opportunities	2.742	3.474	4.254	3.750	1.422	2.846	4.654
Average evaluation	3.386	3.872	4.393	3.906	0.789	3.451	4.362

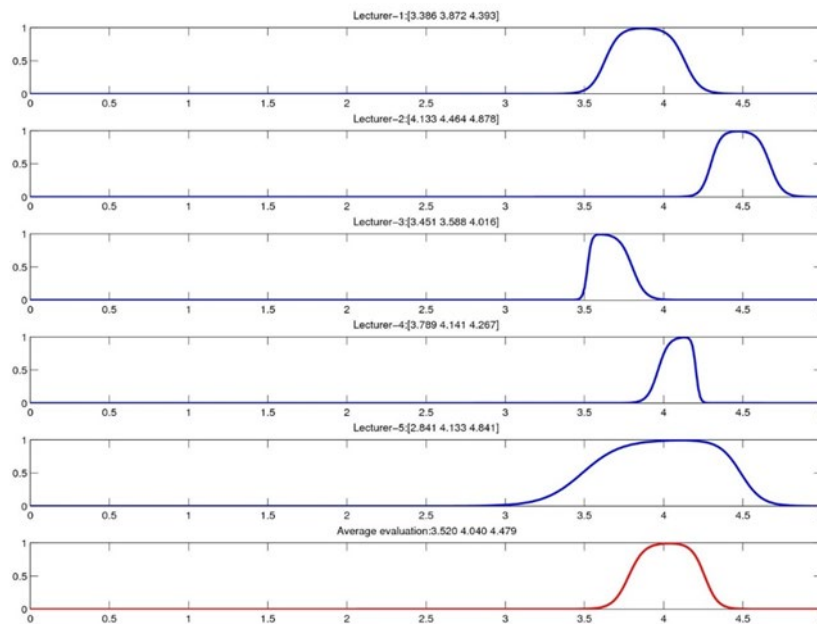


Figure 05. Average evaluation of each lecturer's performance |

7. Conclusion

The evaluation of service quality conducted either by students, peer reviewers or other stakeholders is mainly realized by the application of Likert-scales. Likert-scales are easy-to-use, however, when utilized to judge service quality, there are some shortcomings as they are not able to reflect subjectivity and certain imprecision inherent to personal judgements. A viable alternative technique by giving an answer to the problems arising from the application of Likert-type scales can be a fuzzy rating scale that is rich and expressive enough to find a value fitting approximately the judgement of raters. By applying Dombi's Pliant Inequality Model, the gained responses can be statistically analysed in a convenient way.

The usefulness of fuzzy rating scales has been demonstrated through the application of a fuzzy Likert-scale to evaluate lecturers' performance in the peer review process launched at Budapest University of Technology and Economics, Faculty Economic and Social Sciences. The proposed methodology can express both the uncertainty in the evaluation of the reviewees and the variability of the reviewed colleague's performance during a single lecture in a quantitative way. The computed average evaluations based on the Pliant Inequality Model can foster the Faculty to compare different lecturers'

performance more reliably and to identify best practices as well. Based on the results and due to the simple application, the suggested method may be considered as a new, viable technique. Besides peer evaluation, the application of the presented methodology arises in evaluations where Likert-type scales are applied traditionally, namely e.g. in the case of student satisfaction, job satisfaction, employee satisfaction measurement. |

Acknowledgments

|The presentation of this paper at the 8th ICEEPSY Conference has been supported by Pallas Athéné Domus Animae Foundation. |

References

- |Abdullah, F. (2005), "HEdPERF versus SERVPERF: The quest for ideal measuring instrument of service quality in higher education sector", *Quality Assurance in education*, Vol. 13, No. 4, pp.305-328.
- Abdullah, F. (2006a), "Measuring service quality in higher education: HEdPERF versus SERVPERF", *Marketing Intelligence & Planning*, Vol. 24, No.1, pp.31-47.
- Abdullah, F. (2006b), "The development of HEdPERF: a new measuring instrument of service quality for the higher education sector", *International Journal of Consumer Studies*, Vol. 30, No. 6, pp.569-581.
- Başaran, M.A., Kalaycı, N. and 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*, Vol. 38, No. 10, pp.12565-12568.
- Bayraktaroglu, G. and Ateş, B. (2010), "Testing the superiority and dimensionality of SERVQUAL vs. SERVPERF in higher education", *The Quality Management Journal*, Vol. 17, No. 1, pp.47-59.
- Brochado, A., (2009), "Comparing alternative instruments to measure service quality in higher education", *Quality Assurance in Education*, Vol. 17, No. 2., pp.174-190.
- Calcagni, A. and Lombardi, L. (2014), "Dynamic Fuzzy Rating Tracker (DYFRAT): a novel methodology for modeling real-time dynamic cognitive processes in rating scales", *Applied Soft Computing*, Vol. 24, pp.948-961.
- Carifio, J. and Perla, R.J. (2007), "Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes", *Journal of Social Sciences*, Vol. 3, No. 3, pp.106-116.
- Chen, T.C. (2001), "Applying linguistic decision-making method to deal with service evaluation problems", *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems*, Vol. 9, No. 1, pp.103-114.
- De Battisti, F., Nicolini, G. and Salini, S. (2005), "The Rasch model to measure service quality", *The ICFAI Journal of Services Marketing*, Vol. 3, No. 3, pp.58-80.
- De Battisti, F., Nicolini, G., and Salini, S. (2010), "The Rasch model in customer satisfaction survey data", *Quality Technology & Quantitative Management*, Vol. 7. No. 1, pp.15-34.
- De Oliveira, O.J. and Ferreira, E.C. (2009), "Adaptation and application of the SERVQUAL scale in higher education" paper presented at POMS 20th Annual Conference, 1-4 May 2009, Orlando, Florida USA.
- De Sáa, S.D.L.R., Gil, M.Á., González-Rodríguez, G., López, M.T. and Lubiano, M.A. (2015), "Fuzzy rating scale-based questionnaires and their statistical analysis", *IEEE Transactions on Fuzzy Systems*, Vol. 23, No. 1, pp.111-126.
- Dombi, J. (2008), "Towards a General Class of Operators for Fuzzy Systems", *IEEE T Fuzzy Syst*, Vol. 16, No. 2, pp.477-484.
- Dombi, J. (2009), "Pliant Arithmetics and Pliant Arithmetic Operations", *Acta Polytech Hun*, Vol. 6, No. 5, pp.19-49.

- Douglas, J. and Douglas, A. (2006), "Evaluating teaching quality", *Quality in Higher Education*, Vol. 12 No. 1, pp. 3-12.
- Ghneim, J. (2013), "Fuzzy concept", *Encyclopedia of Psychometrics*, pp.61-68.
- Gil, M.Á. and González-Rodríguez, G. (2012), "Fuzzy vs. Likert scale in statistics" In Trillas, E., Bonissone, P.P., Magdalena, L. and Kacprzyk, J. (Eds.): *Combining experimentation and theory*, Springer, Berlin Heidelberg, pp.407-420.
- 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*, Vol. 27, No. 2, pp. 182-191.
- Gruber, T., Fub, S., Voss, R., and Zikuda, M.G. (2010), "Examining student satisfaction with higher education services - Using a new measurement tool", *International Journal of Public Sector Management*, Vol. 23, No.2, pp.105-123
- Herrera, F., López, E., Mendana, C. and Rodríguez, M.A. (1999), "Solving an assignment-selection problem with verbal information and using genetic algorithms 2.", *European Journal of Operational Research*, Vol. 119, No.2, pp. 326-337.
- Herrera, F., Herrera-Viedma, E. (2000), "Choice functions and mechanisms for linguistic preference relations", *European Journal of Operational Research*, Vol. 120, No. 1, pp.144-161.
- Hesketh, B., Pryor, R., Gleitzman, M. and Hesketh, T. (1988), "Practical applications and psychometric evaluation of a computerized fuzzy graphic rating scale", *Advances in Psychology*, Vol. 56, pp.425-454.
- Jamieson, S. (2004), "Likert scales: how to (ab) use them", *Medical Education*, Vol. 38, No. 12, pp.1217-1218.
- Jonessen, D. H. (2000), "Integrating problem solving into instructional design" In Dempsey, J. and Reiser, R.A.(Eds.), *Trends and issues on instructional design and technology*, Prentice-Hall, Upper Saddle River, NJ
- Kacprzyk, J. (1986), "Towards a 'human-consistent' multistage decision making and control models using fuzzy sets and fuzzy logic", *Fuzzy Sets and Systems*, Vol. 18, No. 3, pp.299-314.
- Lalla, M., Facchinetti, G. and Mastroleo, G. (2005), "Ordinal scales and fuzzy set systems to measure agreement: an application to the evaluation of teaching activity", *Quality & Quantity*, Vol. 38, No. 5, pp.577-601.
- Liu, N., and Carless. D. (2006), "Peer Feedback: The Learning Element of Peer Assessment." *Teaching in Higher Education*, Vol. 11, No. 3, pp. 279–290.
- Liu, R., Cui, L., Zeng, G., Wu, H., Wang, C., Yan, S. and Yan, B. (2015), "Applying the fuzzy SERVQUAL method to measure the service quality in certification and inspection industry", *Applied Soft Computing*, Vol. 26, pp.508-512.
- Lozano, L.M., García-Cueto, E. and Muñoz, J. (2008), "Effect of the number of response categories on the reliability and validity of rating scales", *Methodology*, Vol. 4, No. 2, pp.73-79.
- Lubiano, M.A., de Sáa, S.D.L.R., Montenegro, M., Sinova, B. and Gil, M.Á. (2016.), "Descriptive analysis of responses to items in questionnaires. Why not using a fuzzy rating scale?", *Information Sciences*, 360, pp.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*, Vol. 40, No. 17, pp.7096-7110.
- Sowa, J.F. (2013), "What Is the Source of Fuzziness?", In Seiging, R., Trillas, E., Moraga, C. and Termini, S. (Eds.), *On Fuzziness*, Springer Berlin, Heidelberg, pp.645-652.
- Venkatesan, S. and Fragomeni, S., 2008. Evaluating learning outcomes in PBL using fuzzy logic techniques. In *19th Annual Conference of the Australasian Association for Engineering Education: To Industry and Beyond; Proceedings of the* (p. 534). Institution of Engineers, Australia
- Yousapronpaiboon, K., (2014), "SERVQUAL: Measuring higher education service quality in Thailand", *Procedia-Social and Behavioral Sciences*, 116, pp.1088-1095.
- Yu, C.M., Tsang, H.T. and Chen, K.S. (2016), "Developing a performance evaluation matrix to enhance the learner satisfaction of an e-learning system", *Total Quality Management & Business Excellence*, Published online 19 Sept 2016, <http://dx.doi.org/10.1080/14783363.2016.1233809>.