

# **An Application of Fuzzy Logic to Service Quality Research: A Case of Fitness Service**

**Jose A. Martínez**

Technical University of Cartagena

**Yong Jae Ko**

University of Florida

**Laura Martínez**

Technical University of Cartagena

This research focused on measuring perceived quality in the context of sport and fitness services using a novel approach in sport management: fuzzy logic. Several analytical procedures have been depicted to operate with fuzzy techniques, which may be applied to empirical research by a wide range of researchers in the sport literature, as well as sport managers. This study showed that fuzzy logic is an attractive method to increase the value of the information collected from customers' evaluations. The implemented procedure overcomes the disadvantages of the research focused on the third-person approach, and minimizes the categorization bias and interaction bias derived from the relationship between verbal and numerical labels. An empirical study of two samples of consumers from two fitness centers illustrates the advantages of this method.

Service quality received attention from numerous scholars (e.g., Brady & Cronin, 2001; Ekinci, 2001; Karatepe, Yavas, & Babakus, 2005; Parasuraman, Zeithaml, & Berry, 1985, 1994; Seth, Deshmukh, & Vrat, 2005), and is considered one of the most important issues in sport marketing, for three main reasons: (1) it is a measure that acts as a proxy to calibrate management performance; (2) it is related to the positioning of the organization; and (3) it is a key determinant of ultimate consumer behavior variables such as customer loyalty. In the past decades, several studies have highlighted the importance of this topic in the field of sport marketing and management (Ko & Pastore, 2004; Shonk & Chelladurai, 2008; Tsitskari, Tsiotras, & Tsiotras, 2006; Wakefield & Blodgett, 1994). In response to a call for a more systematic service quality framework, several scholars proposed conceptual models and

---

Martínez and Martínez are with the Dept. of Management and Marketing, Technical University of Cartagena, Cartagena 30203, Spain. Ko is with the Dept. of Tourism, Recreation, and Sport Management, University of Florida, Gainesville, FL.

measurement tools of service quality for sport services (e.g., Kim & Kim, 1995; Ko & Pastore, 2005; Ko, Durant, & Mangiantini, 2008; Wakefield, Blodgett, & Sloan, 1996). However, only a few studies have been conducted to examine the adequacy of the rating scales in the sport management literature. Traditionally, this issue has been a main topic of debate in other disciplines, such as psychology and sociology.

During the first decades of the past century, social science researchers have been searching for an ideal measurement method that minimizes bias and maximizes utility (e.g., Thurstone, 1928; Likert, 1932). Rating scales have been the most used measurement method since Likert (1932) introduced his scale to measure attitudes. However, during the last several decades, researchers have discussed numerous concerns related to how to implement this method (e.g., Cox, 1980; Lissitz & Green, 1975; Preston & Coleman, 2000). Several examples of these debates are: (1) the proper number of response alternatives in the rating scale; (2) the choice of verbal labels to identify the different alternatives; (3) reliability and validity of the scales with different anchors; (4) the effect of incorporating a neutral point; and (5) consumers' preferences among the different response formats. The scholarly discussion of the issues created several alternatives such as semantic-differential scales and other categorical measurement scales with limited response alternatives. There are several excellent references in the literature that review the debate (e.g., Cox, 1980; Hofmans, Theuns, & Mairesse, 2007; Weng, 2004). These studies also demonstrated numerous contradictions that have been derived from empirical research.

In marketing and management research, the most frequently used rating scales are 5 and 7-point Likert-type scales. Numerous scholars have been supportive of using the Likert-type scales (e.g., Cox, 1980; Lissitz & Green, 1975). Likewise, semantic differential scales have also been widely used since the introduction of Prospect Theory (Kahneman & Tversky, 1979) and increased attention on the distinction between positive and negative emotions in evaluation processing (Herzberg, Mausner, & Snyderman, 1959). For example, people think differently when they evaluate a positive outcome than when they evaluate a negative outcome with the same probability of occurrence; losses loom larger than gains, and negative information is more perceptually salient than positively valenced information. Therefore, individuals could be more comfortable evaluating a negative outcome with a negative number than with a positive number.

In terms of maximization of scale utility, it would be theoretically appropriate to consider that consumers give their responses using their preferred scale. Therefore, responses would not be restricted to a categorical scale proposed by the researcher. In this situation, individuals would respond using the "best scale" (i.e., the scale that fits with their preferences) within the context of the question and the answer, and finally using the scale that minimizes their psychological costs (Ferrando, 2003; Weng & Cheng, 2000). We call this "best scale" a "free-scale." This would be, in relativistic terminology, a "first-person approach" for measuring psychological variables (Kilpatrick & Cantril, 1960). This relativistic perspective contrasts with the positivistic "third-person approach" (i.e., when individuals ought to respond to a given scale proposed by the researcher).

Beyond philosophical debates, it seems evident that the first-person approach increases the utility of responses. Some studies have shown how individuals give their responses in a homogeneous rank (Ferrando, 2003), and how they prefer a 10-point scale (Preston & Colman, 2000). Consequently, it would be plausible that individuals use a

similar mental framework to respond. However, it is recommendable to analyze each specific research context (Ferrando, 2003; Hofmans, Theuns, & Van Acker, 2009).

In addition to enhancing utility to avoid categorization bias, another advantage of adopting the first-person approach in marketing and management research is to avoid the interaction effect between the verbal and numerical labels of the rating scale. Note, for example, that cultural differences may yield divergences in the response patterns (Hofmans et al., 2009). Indeed some individuals can consider, for example, “very good” as the final point of a scale, whereas other individuals do not (i.e., the response style may vary depending on the personality of the individuals who respond to the questionnaire; Javaras & Ripley, 2007; Saris & Gallhofer, 2007). As Windschitl and Wells (1996) indicated, numerous studies have shown that there is substantial difference between the subjects’ variability in the numeric interpretations given for words such as *likely*, and phrases such as *reasonable chance* and *close to certain*. In addition, interpretation of verbal expressions differ depending on the context to which these expressions were referring.

Considering the aforementioned arguments, we assume the following premise: data originating from a free-scale are the best or the most accurate representation of individual responses. However, when using rating scales with different anchors than the free-scale, deviations from this ideal situation should be statistically quantifiable. We mean that researchers need to know if a specific rating scale is statistically less or equally useful than the free-scale. In other words, we need to examine if scale invariance holds (i.e., if the measure of the variable of interest does not change if the length of the scale is multiplied by a constant factor, for example, when the free-scale is a 10-point scale and the proposed scale in the questionnaire is a 5-point scale). In addition, we may need to study the relationship between this optimized numerical response and a verbal label associated with this number. Therefore, we will obtain the degree of uncertainty of this verbal response, how individuals use these verbal labels, and how these responses are distributed. Consequently, we will obtain more valuable information regarding consumer perceived quality in sport services, than by using the traditional response formats “strongly disagree-strongly agree,” or “poor quality-good quality” with 5 or 7 response categories. We will achieve this aim by using fuzzy logic.

The purpose of this study was to examine the feasibility of using fuzzy logic in measuring sport consumers’ quality perceptions and provide new useful insights for measurement of service quality in the sport industry. The contribution of this research to sport management literature is twofold: a) to depict a methodology that has not been previously used in the field of sport management and marketing; and b) to analyze the relationship between numerical and verbal responses of consumers who evaluate perceived quality of sport services. In addition, this research applies the fuzzy logic methodology using a disparate perspective from the usual applications (e.g., Lozano & Fuentes, 2003; Nejati, Nejati, & Shafaei, 2009; Tsai, Wu, & Liang, 2008), because fuzzy membership functions are not theoretically, but empirically derived.

## Literature Review

### Perceived Service Quality

The perception of service quality has been extensively studied during the past three decades. Owing to the intangible, heterogeneous, and inseparable nature of services,

service quality has been defined as “the consumer’s judgment about a product’s overall excellence or superiority” (Zeithaml, 1988, p. 3) or “the consumer’s overall impression of the relative inferiority/superiority of the organization and its services” (Bitner & Hubbert, 1994, p. 77).

Several different approaches were proposed to measure service quality in the field of marketing and management (e.g., Brady & Cronin, 2001; Cronin & Taylor, 1992; Dabholkar, Thorpe, & Rentz, 1996; Grönroos, 1982, 1984; Parasuraman, Zeithaml, & Berry, 1985, 1988, 1994; Rust & Oliver, 1994). The common feature of the previously proposed model is that they believed service quality should be conceptualized as a multidimensional construct that is inherently linked to the measurement of consumer quality perceptions. In addition, service quality models offer a framework for understanding what service quality is, as well as how to measure service quality in each proposed conceptualization.

Meanwhile, as Ko, Durrant, and Mangiantini (2008) indicated in their recent work, studies of service quality have also been conducted in various segments of the sport industry, such as professional sport (Kelly & Turley, 2001; Ko, 2005; McDonald, Sutton, & Milne, 1995; Milne & McDonald, 1999; Theodorakis, Kambitsis, Laios, & Koustelios, 2001; Wakefield et al., 1996), fitness programs (Kim & Kim, 1995; Papadimitriou & Karateroliotis, 2000), and recreation and leisure services (Crompton & MacKay, 1989; Crompton, MacKay, & Fesenmaier, 1991; Howat, Absher, Crilley, & Milne, 1996; Ko & Pastore, 2004, 2005; MacKay & Crompton, 1988; Taylor, Sharland, Cronin, & Ballard, 1993; Wright, Duray, & Goodale, 1992). In the Spanish context, there were also relevant studies such as the work of Hernández (2001) and Morales, Hernández, and Blanco (2005). These studies focused on identifying dimensions of quality based on customers’ evaluations and perceptions of the services offered.

Although service quality is a mature area in marketing research, there are still several issues that are a matter of discussion among scholars, such as the multidimensionality problem, the specification of reflective versus formative models, and the use of single versus multiple measures (see Rossiter, 2002). Martínez and Martínez (2010) reviewed the topics and proposed several alternative options for measuring service quality which may help in overcoming the theoretical and methodological shortcomings of the current service quality research. However, the purpose of our study is not to discuss all these concerns, but focus on the analysis of the relationship between numerical and verbal responses of consumers who evaluate the quality of sport services, and introduce fuzzy theory, a new methodological approach that has not been previously used in the field of sport management and marketing. Therefore, our paper focuses on measuring perceived quality in sport services, without going more deeply into the aforementioned controversial issues in service quality research.

## Fuzzy Logic

Fuzzy logic is a form of representation based on the Fuzzy Set Theory, proposed by Zadeh (1965). Fuzzy logic originally came from engineering, and has been mainly used to control systems, with application areas such as air conditioner controllers, digital image processing, or automobile cruise control. However, in the last years, applications have been extended to solve business problems (e.g., Gil-Aluja, 1999; Huang, Chu, & Chiang, 2008).

Fuzzy logic tries to capture the way individuals think, considering the uncertainty of their thoughts. As Rajasekaran and Vijayalakshmi (2003) stated, the uncertainty may arise due to partial information about the problem, or due to information which is not fully reliable, or due to inherent imprecision in the language with which the problem is defined, or due to receipt of information from more than one source about the problem which is conflicting.

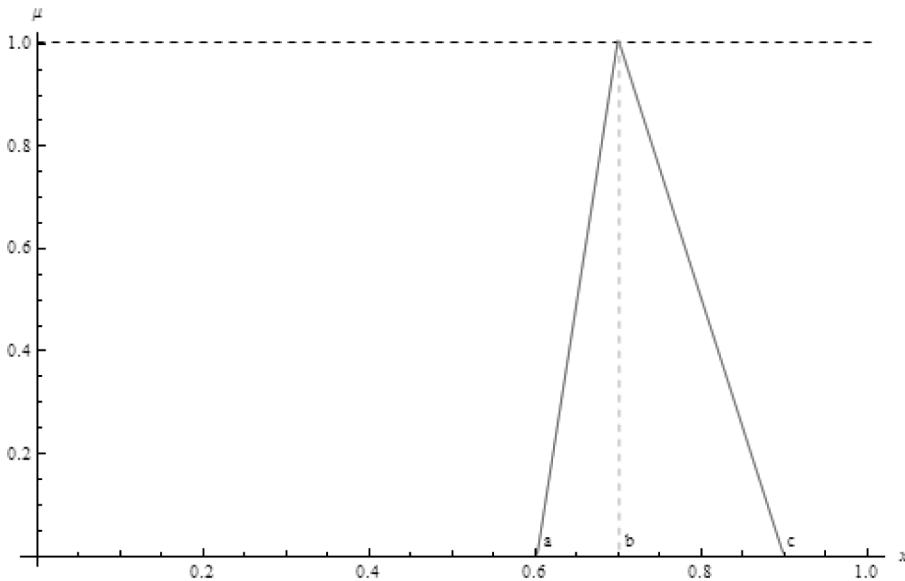
Fuzzy logic is a type of logic that goes beyond the dualistic distinction between true and false propositions. Using fuzzy logic, propositions can be represented with a certain degree of vagueness (i.e., with degree of truthfulness or falseness). For example, the sentence "today is a sunny day," may be 100% true if there are no clouds in the sky, 80% true if there are a few clouds in the sky, 50% true if it is a foggy day, and 0% true if the clouds completely cover the sky (Lozano & Fuentes, 2003). Using a similar reasoning, the perception of quality a customer may have toward a sport center service may be represented by using a membership function that relates each linguistic response with a set of values enclosed in a  $[0,1]$  interval. For example, when a customer claims that the service quality of a sport center is "good," this linguistic term has an uncertain meaning, because the word "good" can have a heterogeneous meaning for different customers depending on their personality, culture, or research context. If a numerical value is assigned to this linguistic term in a universal  $[0,1]$  scale, then a set of probable values forming a fuzzy number could be symbolized (Zwick & Wallsten, 1989). In this specific case, it would be plausible to propose a fuzzy number, ranked, for example, from 0.6 to 0.9. These ranks of values have an associated probability. The membership function is usually determined a priori, by assigning a fuzzy number to a linguistic term previously proposed by the researcher (i.e., the third-person perspective; see Chen, 2000; Liu & Wang, 2008, for generalized applications, and see Lozano & Fuentes, 2003; Nejati et al., 2009; Tsai et al., 2008; Tseng, 2009, for applications to service quality research). An illustrative example of this way to proceed is the definition of fuzzy numbers for Likert scales from 1 to 5, with verbal labels from "strongly disagree" to "strongly agree."

The perspective that we adopted in this research is different. The membership function is derived from the empirical data, by relating the numerical values that reflect the customer's subjective perception of quality and the linguistic terms used to define this attitude (first-person approach). Therefore, there is no a priori assumption about the distribution of fuzzy values.

The applications of fuzzy set theory in this study, following the notation of Tsai et al. (2008), are elaborated as follows:

In a universe of discourse of  $X$ , a fuzzy subset  $\tilde{A}$  of  $X$  is characterized by a membership function  $\mu_{\tilde{A}}(x)$  which associates a real number in the interval  $[0, 1]$  with each element  $x$  in  $X$ , to represent the "grade of membership" of  $x$  in  $\tilde{A}$ . A fuzzy number in real line is a triangular fuzzy number  $\tilde{A}$  if its membership function  $\mu_{\tilde{A}}(x) : \mathbb{R} \rightarrow [0,1]$  is:

$$\mu_{\tilde{A}}(x) = \begin{cases} \frac{(x-a)}{(b-a)} & , \quad a \leq x \leq b \\ \frac{(x-c)}{(b-c)} & , \quad b \leq x \leq c \\ 0 & , \quad \text{otherwise} \end{cases}$$



**Figure 1** — The triangular membership function of the fuzzy number.

which can be expressed as the triplet  $\tilde{A} = (a, b, c)$  with  $a \leq b \leq c$ . The “mean” of the fuzzy number is represented by  $b$ , which corresponds to the peak of its membership function:  $\mu_{\tilde{A}}(x) = 1$ , while  $a$  and  $c$  denote a minimum value on the left representation and a maximum value on the right representation of a fuzzy number, respectively. The membership value increases gradually (linearly) from  $a$  to  $b$  and decreases from  $b$  to  $c$ . Figure 1 illustrates the triangular membership function of the fuzzy number.

We have chosen a discrete triangular function due to the easiness to operate with this type of function. Other possible types of functions can be consulted in Rajasekaran and Vijayalakshmi (2003).

The numerical response of each individual can be transformed into a fuzzy number empirically derived if they provide their responses using an interval. For example, a customer can judge his/her quality perception with a value of 0.5, but with some degree of uncertainty represented in the interval (0.4, 0.6). If other customers respond using a unique value (i.e., with a minimum uncertainty about their quality perception), then the number is a crisp number (i.e., a nonfuzzy number).

In addition, if customers evaluate their quality perceptions using linguistic terms, a fuzzy relationship can be established between the verbal labels and the numbers (crisp numbers) associated with them. For example, the numerical evaluations associated with the term “good quality” are multinominally distributed in the sample, because the different numerical values of this word have an associated probability. The expected value of the distribution is the maximum of the membership function:  $b'$ , while  $a'$  and  $c'$  are the minimum and maximum of the distribution. When there is no variation in the probability distribution of numeric responses for a specific verbal term, then the fuzzy number is a crisp number.

Formally, in a given set of  $m$  linguistic terms  $S_\theta = \{S_i\}_{i=1\dots m}$ , then there is a set of  $m$  triangular fuzzy numbers  $\tilde{A}_\theta = \{\tilde{A}_i\}_{i=1\dots m}$ ,  $\tilde{A}_i = \{a_i', b_i', c_i'\}$  and  $\mu_{\tilde{A}_i}(x)$  is:

$$\mu_{\tilde{A}_i}(x) = \begin{cases} \frac{(x - a_i')}{(b_i' - a_i')} & , \quad a_i' \leq x \leq b_i' \\ \frac{(x - c_i')}{(b_i' - c_i')} & , \quad b_i' \leq x \leq c_i' \\ 0 & , \quad \text{otherwise} \end{cases}$$

being  $b_i' = \sum p_{ij} x_{ij}$  where  $p_j$  is the probability of an individual  $j$  respond the value  $x_j$  for each linguistic term  $i \in m$ . Finally,  $a_i' = \min(x_{ij})$  y  $c_i' = \max(x_{ij})$ .

Fuzzy numbers can be compared in function of their degree of similarity. Although several similarity measures have been proposed (e.g., Bojadziev & Bojadziev, 1996; Guha & Chakraborty, 2010; Zwick, Carlstein, & Budescu, 1987), we used a straightforward method, computing the intersection between two fuzzy numbers (Zwick et al., 1987, p. 227),  $a_\cap = \tilde{A} \cap \tilde{B}$ , and the percentage of the area that take up this intersection in relation with the sum of the areas of both fuzzy numbers  $a_\cup = \tilde{A} \cup \tilde{B}$ . Therefore, two fuzzy numbers will be totally disparate if  $a_\cap/a_\cup = 0$ , and totally similar if  $a_\cap/a_\cup = 1$ .

## Methods

### Data Collection and Measures

We collected data from two pseudo-random samples of customers from two main sport centers of a Spanish city. Data were collected during the spring of 2008 by employing personal interviews. Business and marketing students were previously trained for this task. The students were placed in the gate of the sport center and they asked each user reasons of abandoning the center. They had to survey customers during one hour per day. In addition, they repeated this task during several days and during different hours, to avoid selection bias, and to guarantee representativeness of the research sample. Refusal rate was about 25% with lack of time being the main reason to refuse. Research participants had to evaluate service quality using a linguistic term and then a numeric value; "*How would you evaluate the quality of this sport center? Please, respond using a linguistic term.*" And, "*how would you evaluate the quality of this sport center using a numeric value?*" When they assigned a numeric value, they had to indicate in which scale that value had meaning. This was taken as the free-scale. They also could indicate their degree of uncertainty associated with that numeric value if necessary. Then, after two questions regarding other variables, the interviewer returned to inquire about service quality by using the following expression: "*You have indicated that your service quality perception was XXX (the interviewer used the same linguistic term that the consumer had used previously), so how would you represent that evaluation on the following scales?*" (semantic differential scale ranged from -3 to +3, and Likert-type scale ranged from 1 to 5 and from 1 to 7). Thus, by separating the questions about service quality into groups, we tried to minimize the possibility of appearance of dependency between numerical responses. We wanted to get a conditional



independent model (see Bollen, 2002; Martínez & Ruiz, in press), where the four considered scales reflect variations in the perceived quality latent variable. This is the common assumption in the reflective view about measurement (Bollen, 2002). Dependency between numerical responses would require a different statistical treatment of data (Bollen & Medrano, 1998).

In addition, research participants had to evaluate their satisfaction with the sport center using also a linguistic term and then a numeric value. The procedure was similar: *"How would you evaluate your satisfaction with this sport center? Please, respond using a linguistic term."* And, *"how would you evaluate your satisfaction with this sport center using a numeric value?"* Again, when they assigned a numeric value, they had to indicate in which scale that value had meaning. This was taken as the free-scale. They could also indicate their degree of uncertainty associated with that numeric value if necessary.

Therefore, we measured service quality and customer satisfaction using one single item, in line with the methodological perspectives of Bergkvist and Rossiter (2007), Drolet and Morrison (2001), Hayduk and Glaser (2000), or Hayduk, Pazderka-Robinson, Cummings, Boadu, Verbeek, and Perks (2007). Other authors such as Babakus and Boller (1992), Bolton and Drew (1991), and Martínez and Martínez (2009) have also used one single measure of service quality. In addition, Rossiter (2002) acknowledged that if quality perception is considered as a concrete attribute (such as an overall assessment of the service), one single item should be enough to correctly measure this variable.

The first sample was composed of 116 individuals of the most popular, publicly managed sport center of the city. Sixty-nine percent of the respondents were male, and 31% were female. The average length of association with the center and the service experience was 31 months. About 40% of the respondents were users of the gymnasium, in accordance with the real percentage of gymnasium users of the center. In addition, other activities were also represented, as fitness, aerobic, martial arts, etc. The study was replicated in a second sample of 98 consumers of a private sport center located in the same city. Sixty percent of the respondents were male, and 40% were female. The mean of the service experience was 40 months. About 70% of the respondents were users of the heated swimming pool. Again, this was representative of the real number of users using this facility. Consumers preferred to give their responses regarding service quality and satisfaction in continuous scales: from 1 to 10, and from 0 to 10. The percentage of responses in the first sample was 52% from 1 to 10 and 48% from 0 to 10 (in the case of both service quality and satisfaction). In the second sample, the percentages were 78% and 22%, respectively (in the case of service quality), and 74% and 24%, respectively, in the case of satisfaction.<sup>1</sup> We say "continuous" because they often used decimal numbers to evaluate service quality. All these responses were transformed to a unit [0, 1] interval. This was the free-scale. The remaining data derived from the other rating scales were also transformed to a unit interval.

## Results

First, we applied *D*-test (Martínez & Ruiz, in press) to analyze scale invariance. *D*-test is a statistical test based on symbolic entropy. Entropy is a measure of uncertainty associated to a random variable (Shannon, 1948), and *D*-test is based



on the definition of Shannon's entropy between different symbols. The hypothesis of this test is that scale invariance holds between pairs of scales. *D*-test is based on the following theorem (Martínez & Ruiz, in press):

Let *P* be a population of cardinality *N* such that each individual  $e \in P$  evaluate a fixed item. Assume that each individual gives its valuation in scale *A* and in scale *B*. Denote by *h*(*S*), *h*(*S*,*A*) and *h*(*S*,*B*) the total symbolic entropy of the valuation, of the valuation with scale *A* and of valuation with scale *B* respectively, as defined in (3). If the valuation with scale *A* does not differ with the valuation with scale *B*, then

$$D(k) = 4N[h(S) - h(S, A) - h(S, B) - \ln(\frac{1}{2})]$$

is asymptotically  $\chi^2_{k-1}$  distributed.  
Let  $\alpha$  be a real number with  $0 \leq \alpha \leq 1$ . Let  $\chi^2_\alpha$  be such that

$$P(\chi^2_{k-1} > \chi^2_\alpha) = \alpha.$$

Then the decision rule in the application of the *D* test at a 100(1 -  $\alpha$ )% confidence level is:

If  $0 \leq D(k) \leq \chi^2_\alpha$  Accept  $H_0$

Otherwise Reject  $H_0$

Results are shown in Table 1. Based on the result, we rejected the hypothesis of scale invariance for the comparison between the free-scale and Likert-type scale (ranged from 1 to 5). However, we cannot reject the hypothesis in the remaining comparisons. Consequently, scale invariance holds when responses are taken in a 7-point scale (Likert from 1 to 7 and semantic differential from -3 to +3).

Next, we analyzed the fuzzy numbers associated with the linguistic terms used by the customers in the first sample. Following the procedure previously depicted, we obtained the fuzzy numbers that are showed in Table 2. Numeric values come from the free-scale. Likewise, Figure 2 shows the oft used linguistic terms.

The procedure implemented to create the fuzzy numbers is very sensitive to outliers. This fact can be viewed in the case of the fuzzy number that represents "good quality," where an individual responded with a very low numeric value, so

**Table 1 D-Test**

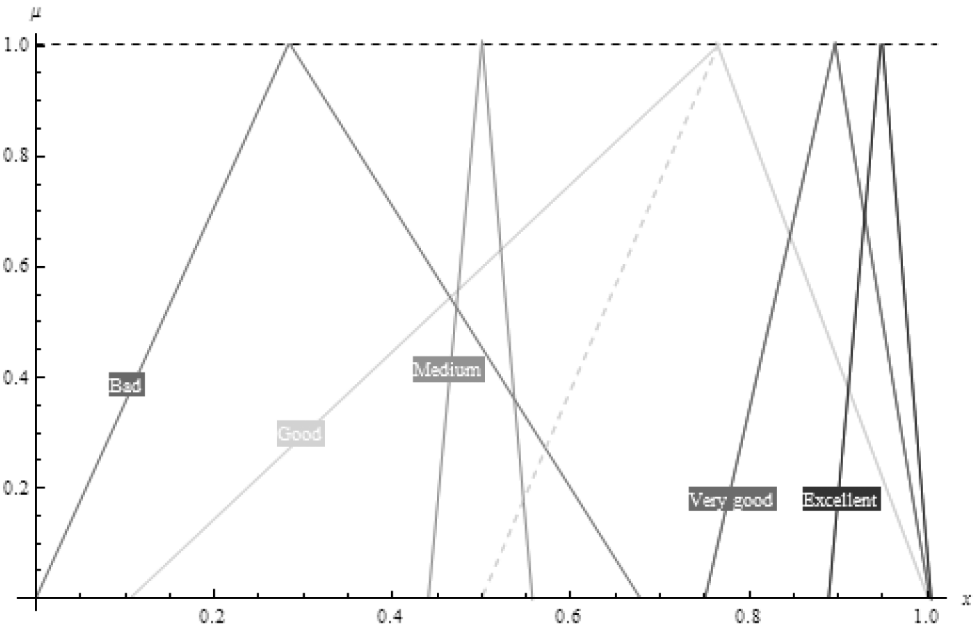
Comparison <sup>a</sup>	<i>D</i> -test	Subintervals ( <i>k</i> )	Cut-off values Chi-square (95%)	Scale invariance
A-B	0.94	7	12.59	Yes
A-C	14.88	5	9.48	No
A-D	6.96	7	12.59	Yes
Replication				
A-B	3.37	7	12.59	Yes
A-C	12.61	5	9.48	No
A-D	9.61	7	12.59	Yes

<sup>a</sup> Free-scale (A) Semantic differential from -3 a +3; (B) Likert type from 1 to 5; (C) Likert type from 1 to 7 (D)

**Table 2    Distribution of Responses for the First Sample (Service Quality). Linguistic and Numerical Relationships.**

Description		Fuzzy numbers			
Linguistic terms	%	a'	b'	c'	Linguistic evaluation
Good	52.17	0.111	0.768	1.000	+
Very good	12.17	0.750	0.895	1.000	+
Bad	8.69	0.000	0.268	0.667	-
Excellent	6.08	0.889	0.977	1.000	+
Medium	5.21	0.444	0.509	0.556	Neutral
Needs to improve	3.47	0.333	0.432	0.550	-
Normal	3.47	0.556	0.639	0.800	Neutral
Acceptable	2.60	0.556	0.644	0.778	+
Insufficient	1.73	0.100	0.275	0.450	-
Right*	1.73		0.770		+
The best*	0.86		1.000		+
Notable*	0.86		0.800		+
Sufficient*	0.86		0.500		+

\* The variance of the distribution of responses is 0, so they are crisp numbers



**Figure 2** — The oft used linguistic terms.

the form of the triangular fuzzy number is severely dependent of this single value. As Figure 2 shows, if we drop this value, the fuzzy number noticeably changes (green dashed line). Therefore, we believe it is convenient to avoid outliers in the subsequent analysis as long as there is a reasonable sample size for each number. Recall that in the case of “good quality,” we have a percentage of responses above 50%, so the detection of the outliers is based on perception. However, when the percentage of responses is very low, it would be dangerous to label a case as an outlier.

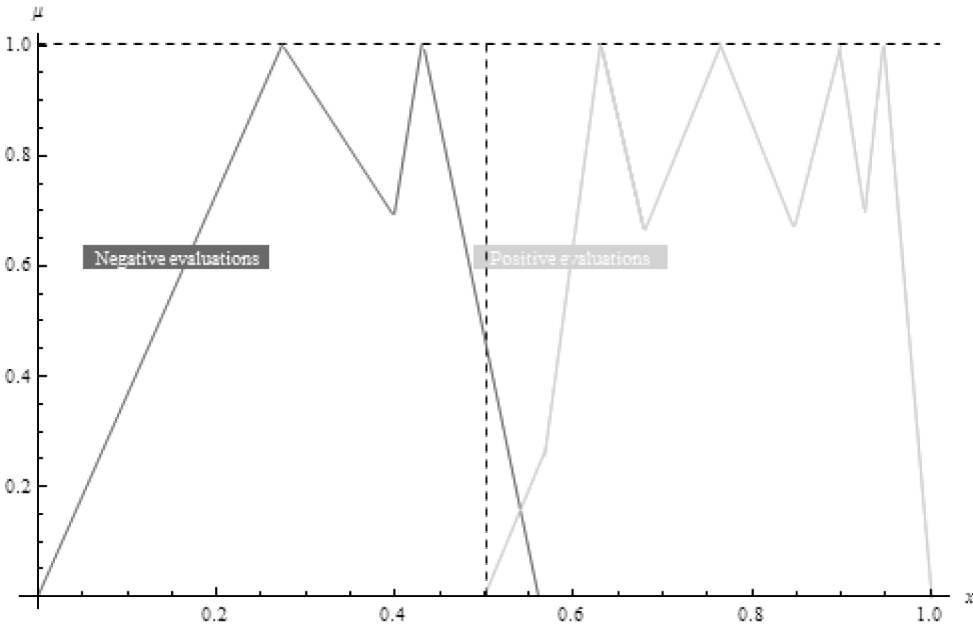
A first relevant conclusion derived from the fuzzy numbers is the following: numerical values from 0.7 to 0.85 have an associated probability to “good quality” higher than the associated probability to “very good quality.” This means that “good quality” may reflect higher numerical values than “very good quality.”

Once the fuzzy numbers have been created, we can operate with them using easy arithmetic rules. For example, we may create two new fuzzy numbers  $\tilde{P}$  and  $\tilde{N}$  representing positive and negative evaluations of the service. To achieve this aim, we classified the linguistic terms as positive, negative, and neutral (Table 2). The computation of  $\tilde{P}$  and  $\tilde{N}$  is achieved as follows:

$$\tilde{P} = \cup \tilde{A}_i(+)$$

$$\tilde{N} = \cup \tilde{A}_i(-)$$

Figure 3 show the values of the membership function for the two new fuzzy numbers. It is noticeable that the degree of membership of a numeric value ranging from 0.5 to 0.55 is higher for the fuzzy number that represents negative verbal evaluations. Consequently, numerical values enclosed in that numerical interval



**Figure 3** — Values of the membership function for the two new fuzzy numbers.

do not necessary mean a positive perceived service quality. However, we must be very prudent with this affirmation because of the low percentage of responses with negative evaluations.

In addition, we can conduct a heuristic analysis to study the difference between the verbal response and the different numerical responses that correspond to the four considered scales. This is a form of studying the interaction between verbal labels and the categorical rating scales. Given that the results of *D*-test support the scale invariance for the scales with seven response categories, it will be expected that this result will be similar after conducting a fuzzy analysis if there is no interaction.

Therefore, we used the geometrical interpretation about the union and intersection of fuzzy numbers explained previously. This comparison was achieved using the three verbal labels with higher percentage of responses: "good," "very good," "bad." Fuzzy numbers are shown in Table 3.

Results of the similarity analysis are shown in Table 4.

**Table 3** Distribution of Responses for the Four Scales for the First Sample. Linguistic and Numerical Relationships for the Three Main Verbal Terms.

Description		Fuzzy numbers			
Linguistic terms	%	a'	b'	c'	Scale
Good	52.17	0.500	0.768	1.000	Free-scale
		0.500	0.808	1.000	Semantic differential
		0.250	0.671	1.000	Likert from 1 to 5
		0.333	0.733	1.000	Likert from 1 to 7
Very good	12.17	0.750	0.895	1.000	Free-scale
		0.833	0.917	1.000	Semantic differential
		0.750	0.885	1.000	Likert from 1 to 5
		0.833	0.923	1.000	Likert from 1 to 7
Bad	8.69	0.000	0.268	0.667	Free-scale
		0.000	0.222	0.500	Semantic differential
		0.000	0.250	0.500	Likert from 1 to 5
		0.000	0.317	0.667	Likert from 1 to 7

**Table 4** Similarity Between Linguistic Terms for the First Sample

	A-B			A-C			A-D		
	$a_{\cap}$	$a_{\cup}$	$a_{\cap}/a_{\cup}$	$a_{\cap}$	$a_{\cup}$	$a_{\cap}/a_{\cup}$	$a_{\cap}$	$a_{\cup}$	$a_{\cap}/a_{\cup}$
Good	0.207	0.411	0.50	0.231	0.341	0.67	0.22	0.25	0.88
Very good	0.13	0.135	0.96	0.073	0.147	0.50	0.079	0.134	0.58
Bad	0.237	0.317	0.74	0.313	0.356	0.88	0.25	0.354	0.70

Free-scale (A); Likert from 1 to 5 (B); Likert from 1 to 7 (C); Semantic differential from -3 to +3 (D).

For the term “good quality,” results are consistent with findings derived from *D*-test. When individuals believe that service quality is “good,” the fuzzy numbers of the invariant scales are much more concordant with the fuzzy number of the free-scale than the noninvariant scale. Similarity index is 67% and 88% for the formers, and only 50% for the latter. However, results are not equally clear for the remaining verbal terms: “very good” and “bad,” but the problem with these two terms is again the low percentage of responses, so the reliability of these results is seriously questioned.

The fuzzy logic approach can be extended to measure service quality dimensions or subdimensions, or other consumer variables such as customer satisfaction. As we collected data about customer satisfaction, we conducted additional analyses. The procedure we propose for that purpose is to achieve a comparative analysis between triangular fuzzy numbers of both variables: quality and satisfaction. Then, we could examine if the meaning of linguistic terms is analogous. In addition, we may sum triangular fuzzy numbers (Bojadziev & Bojadziev, 1996) to build a new fuzzy number derived from the prior two. As Yim, Tse, and Chan (2008) postulated, customers develop affectionate bonds with firms through their repeated experience of high levels of service quality and satisfaction. This enhances customer loyalty. Therefore, to the extent that the new triangular fuzzy number derived from the sum of the fuzzy quality and satisfaction numbers is high, then affective commitment with the sport center increases.

Following the same procedure than before, we obtained the fuzzy numbers that are showed in Table 5.

Comparing these results with the responses regarding service quality, we observe that the verbal pattern is very alike. We again focused our analysis on the term “good” that is the most representative. The similarity between both fuzzy numbers of quality and satisfaction is really high:  $a_{\cap}/a_{\cup} = 88\%$ , so there is a high concordance between verbal and numeric terms for both variables. This similarity index means that the fuzzy weighted sum (or normalized sum) between both numbers will be almost the same number. As expected the new fuzzy number was very similar to each separate number: (0.5, 0.78, 1).

We then repeated the same analyses in the replication sample. Table 6 shows the distribution of responses.

Comparing these results with the first sample, we observe that the verbal response pattern is very alike. Four of the five most used terms in the first sample are the most used in the second sample. Again, the term “good quality” is noticeably the main verbal label.

**Table 5    Distribution of Responses for the First Sample (Customer Satisfaction). Main Linguistic and Numerical Relationships**

Description		Fuzzy numbers			Linguistic evaluation
Linguistic terms	%	a'	b'	c'	
Good	52.59	0.50	0.79	1.00	+
Very good	9.48	0.75	0.91	1.00	+
Bad	6.90	0.00	0.25	0.50	-
Satisfied	6.03	0.44	0.74	1.00	+
Excellent	5.17	0.89	0.95	1.00	+

**Table 6** Distribution of Responses for the Second Sample (Service Quality). Linguistic and Numerical Relationships.

Description		Fuzzy numbers			Linguistic evaluation
Linguistic terms	%	a'	b'	c'	
Good	56.70	0.556	0.743	1.000	+
Medium	15.46	0.222	0.510	0.778	Neutral
Very good	7.21	0.778	0.869	1.000	+
Bad	4.12	0.222	0.250	0.333	-
Needs to improve	3.09	0.556	0.624	0.667	-
Acceptable	2.06	0.556	0.628	0.700	+
Right	2.06	0.600	0.667	0.667	+
Excellent*	1.03		1.000		+
Insufficient*	1.03		0.500		-
Really good*	1.03		0.889		+
Pitiful*	1.03		0.222		-
Sufficient*	1.03		0.500		+
Efective*	1.03		0.889		+
Really bad*	1.03		0.100		-
Appropriate*	1.03		0.778		+
Nice*	1.03		0.800		+

\* The variance of the distribution of responses is 0, so they are crisp numbers

In Figure 4, we illustrate this comparison between both samples. Results are practically identical for the terms “good quality” and “very good quality,” while they differ for “medium quality” and “bad quality.” For these latter two terms, however, the number of responses is very low, so we must be prudent in our interpretation. In this second sample, there is again a noticeable point to comment regarding the overlapping of fuzzy numbers. In the first sample, we compared the terms “good quality” and “very good quality,” and in this second sample, we may compare the terms “medium quality” and “good quality,” because they are the most used terms. Therefore, numerical values between 0.55 and 0.64 have an associated probability to the term “medium quality,” higher than the associated probability of the term “good quality.” Consequently, numerical evaluations that we could consider as positive, do not mean “good” for the customer. Likewise, the terms “medium quality” and “good quality” have a high fuzziness. This means that the heterogeneity of their meaning is high. If we focus on the term “good quality,” that is the main term in both samples, the membership function practically covers the entire rank of numerical evaluations that could be considered as positive.

We again conducted an analysis to study the difference between the verbal response and the different numerical responses that correspond to the four considered scales. Because of the low percentage of responses for the majority of the verbal terms, we only focused on the term “good quality.” Tables 7 and 8 show the results of this comparison.



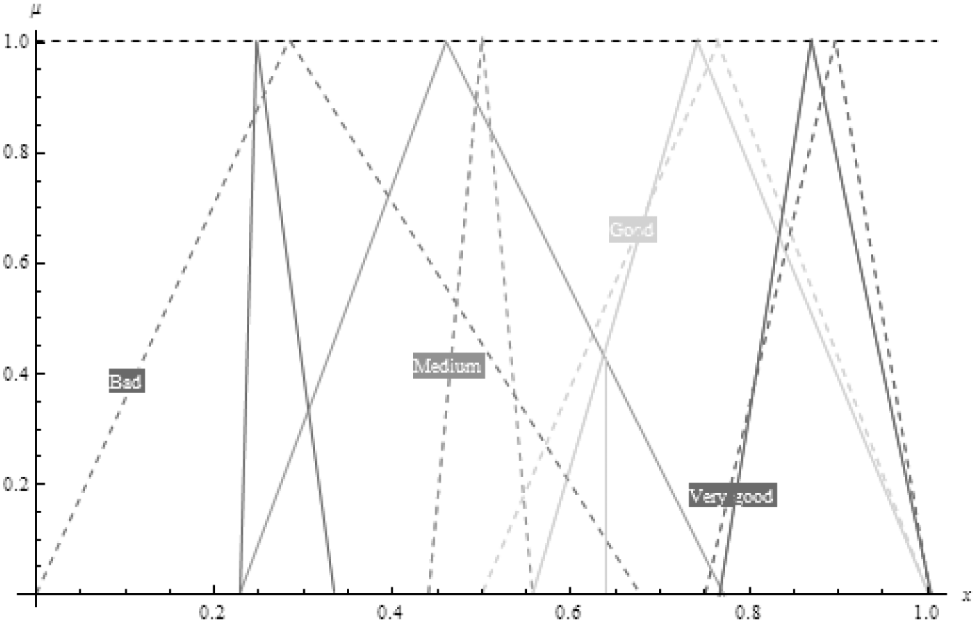


Figure 4 — Comparison between samples.

Table 7 Distribution of Responses for the Four Scales for the Second Sample (Service Quality). Linguistic and Numerical Relationships for the Main Verbal Term

Description		Fuzzy numbers			
Linguistic terms	%	a'	b'	c'	Scale
Good	56.70	0.556	0.743	1.000	Free-scale
		0.500	0.803	1.000	Semantic differential
		0.250	0.655	1.000	Likert from 1 to 5
		0.333	0.709	1.000	Likert from 1 to 7

Table 8 Similarity Between Linguistic Terms for the Second Sample (Service Quality)

	A-B			A-C			A-D		
	a <sub>∩</sub>	a <sub>∪</sub>	a <sub>∩</sub> /a <sub>∪</sub>	a <sub>∩</sub>	a <sub>∪</sub>	a <sub>∩</sub> /a <sub>∪</sub>	a <sub>∩</sub>	a <sub>∪</sub>	a <sub>∩</sub> /a <sub>∪</sub>
Good	0.188	0.413	0.45	0.205	0.352	0.58	0.253	0.276	0.92

Free-scale (A); Likert from 1 to 5 (B); Likert from 1 to 7 (C); Semantic differential from -3 to +3 (D)

**Table 9** Distribution of Responses for the Second Sample (Customer Satisfaction). Main Linguistic and Numerical Relationships.

Description		Fuzzy numbers			Linguistic evaluation
Linguistic terms	%	a'	b'	c'	
Good	43.88	0.44	0.72	1.00	+
Medium	12.24	0.33	0.56	1.00	Neutral
Right	10.20	0.66	0.75	0.89	+
Very good	6.12	0.78	0.90	1.00	+
Satisfied	4.08	0.66	0.72	0.78	+

As in the first sample, the 5-point Likert scale has a more disparate behavior. Similarity index for the 7-point Likert scale is lower in this second sample (58%) than in the first sample (67%), but the results for the semantic differential scale are very alike: 91% and 88%, respectively.

Again, we obtained the fuzzy numbers for customer satisfaction that are outlined in Table 9. Comparing these results with the responses regarding service quality in this second sample, we observed that the verbal pattern was similar. We again focused our analysis on the term “good” that is the most representative. The similarity between both fuzzy numbers of quality and satisfaction is not as high as in the first sample:  $a_{\cap}/a_{\cup} = 71\%$ , because the satisfaction fuzzy numbers are more heterogeneous than the quality fuzzy number. The difference is due to the minimum extreme of the triangular fuzzy number, which in the case of satisfaction, could be associated to a number below the average value of the scale (from 0.5 to 0.44). The fuzzy weighted sum between both numbers was: (0.49, 0.73, 1).

Finally, it is highly noticeable that only four customers expressed their perceived quality and satisfaction indicating an uncertainty numerical interval. This percentage is obviously negligible, so we have omitted the fuzzy analysis that could arise from this.

### Implications

In this research, we have focused on the measure of perceived quality in sport services using an underutilized method in sport management: the fuzzy logic approach. We have depicted several straightforward procedures to operate with fuzzy techniques, which may be applied to empirical research by a wide range of researchers and sport managers. The main limitation of our study is the low sizes of the two samples, because higher samples would have allowed us to create more consistent and reliable fuzzy numbers (see Apendix).

Despite this shortcoming, this study still makes a valuable contribution. First, consumers of sport services prefer to evaluate perceived quality using scales ranging from 1 to 10 or from 0 to 10. However, results derived from invariance analysis show that 7-point scales are invariant but 5-point scales are not. This is a relevant issue for applied research and meta-analysis, because 5-point scales yield biased outcomes. Therefore, studies in the sport management field, such as Ko and Pastore

(2005) or Ko et al. (2008) cannot be questioned regarding scale invariance, because they used a 7-point scale, while results of other relevant research, such as Getz, O'Neill, and Carlsen's (2001), Hernández's (2001), and Morales et al.'s (2005) study can be questioned, because they used a 5-point scale.

Second, fuzzy logic analysis shows that there is not a "perfect" homogenous relationship between linguistic terms and numerical values. Fuzzy numbers that reflect positive and negative evaluations indicate that, in the first sample, numerical values ranging from 0.5 to 0.55 (i.e., associated to a positive interpretation) have a higher probability to be linked to a verbal term with negative connotations. In the second sample, numerical values ranging from 0.55 to 0.64 have a higher probability to be linked to a neutral verbal term than to a verbal term with positive connotations. Therefore, numerical evaluations that we could interpret as positive may not really mean a positive quality perception for the customer. This result is very important, because it proves that it is not enough to know the customer's quality perception derived from a numeric value if the researcher does not know what this number actually means for this customer.

Third, there is a great degree of heterogeneity of linguistic terms that customers freely use to express their quality perception. However, the term "good quality" is, undoubtedly, the most common. This is the term that permits a reliable interpretation because of the high percentage of responses. "Good quality" is a high fuzziness term (i.e., it means different things for disparate individuals). This fact is supported in other studies that demonstrate the heterogeneity of the meaning of linguistic terms (Wallsten, Budescu, & Zwick, 1993). Therefore, it would be expected that the commonly used rating scales that relate linguistic labels with numerical labels yield distorted results, because of the effect of interaction. However, we have shown that this interaction effect is not so important for invariant scales, because there is much more agreement between the fuzzy numbers of these invariant scales and the fuzzy numbers of the free-scale. Consequently, it seems that when scale invariance holds, there is higher homogeneity in the relationship between numerical and verbal responses.

Fourth, and unexpectedly, customers evaluate service quality without uncertainty in their numerical responses. This means that individuals are able to express their thoughts with high numerical precision. Customers know perfectly what it means, for example, that their quality perception is "good," because of the use of a crisp number to evaluate service quality. Hence, there is no individual uncertainty susceptible to be modeled. This contradicts the commonly used third-person approach (e.g., Lozano & Fuentes, 2003; Nejati et al., 2009; Tsai et al., 2006; Tseng, 2009), and favors our first-person approach, because we allow customers to assign a meaning to their evaluations, and then we aggregate these meanings by creating a fuzzy number. Therefore, studies that use fuzzy logic under a third-person approach may yield distorted results, because the *a priori* fuzzy numbers designed by the researcher can be quite far from the fuzzy number empirically derived.

Obviously, the use of this methodology is much more difficult when implementing comprehensive measurement models of service quality (i.e., using multi-item scales for measuring dimensions and subdimensions of service quality; e.g., Brady & Cronin, 2001; Ko & Pastore, 2005; Martínez & Martínez, 2007). Recall that applications of fuzzy logic to multi-item scales can be encountered in the literature (e.g., Nejati et al., 2009; Tseng, 2009), but all these research use the

third-person approach (i.e., fuzzy numbers are theoretically derived). We stress that this approach may distort results because the empirical meaning of linguistic terms can be different from the *a priori* expected.

Therefore, the advocacy of authors such as Martínez and Martínez (2010) or Rossiter (2002), of simplifying service quality measures could make the application of fuzzy logic to service quality measurement easier. Measuring service quality attributes (dimensions or subdimensions) by using one single item would make it easier to analyze the relationships between verbal and numerical responses to each specific service quality attribute. An example of operating with more variables has been demonstrated, analyzing a customer satisfaction measure. This can be extended to measures of service quality attributes or other attitudinal variables, as truth or corporate image. Therefore, researchers can create new fuzzy numbers by aggregation, using the weighted sum procedure. This could lead researchers to achieve a profile of the meaning of linguistic terms, adding all attitudinal variables (service quality attributes, for example).

Our analysis of customer satisfaction has shown that linguistic terms used are very alike. Again the term "good" is the most represented, which shows high concordance with the term "good quality." However, the replication analysis demonstrated that good satisfaction may be associated to a number with negative connotations (below 0.5), so it seems clear that both evaluations (quality and satisfaction) are based on some disparate judgments.

We advocate for the use of both linguistic and numeric terms for analyzing service quality when the characteristics of the research allow for that. Recall that numeric measures do not always provide the adequate means for assessing psychological uncertainty (Windschitl & Wells, 1996), and individuals may prefer to give their responses in any of the forms: verbal or numerical (Weinstein & Diefenbach, 1997)<sup>2</sup>.

Finally, upon acknowledging the shortcomings of this research, we may conclude that the fuzzy logic approach that we propose is an appropriate method to increase the value of the information collected from customers' evaluations. The procedures we have implemented overcome the disadvantages of the third-person approach, and minimize the categorization bias and interaction bias derived from the relationship between verbal and numerical labels. Measuring and analyzing perceived quality of sport services in this way is relatively straightforward from an operative viewpoint. We encourage the use of this methodology in sport management research.

## Notes

1. A minimum percentage (2%) of customers preferred the 1–5 scale.
2. In a small and independent experiment conducted by us, customers of sport services preferred to give their evaluations of service quality in the following form: numerically (24.2%), verbally (25.8%), and either of the two forms (50%). Sample size was 62 customers only, but the obtained percentages are illustrative of the extent to which verbal judgments can be well received by respondents.

## References

- Babakus, E., & Boller, G.W. (1992). An empirical assessment of the SERVQUAL scale. *Journal of Business Research*, 24, 235–268.

- Bergkvist, L., & Rossiter, J.R. (2007). The predictive validity of multiple-item vs. single-item measures of the same constructs. *JMR, Journal of Marketing Research*, 44(2), 175–184.
- Bojadziev, G., & Bojadziev, M. (1996). *Fuzzy sets, fuzzy logic, applications*. Singapore: World Scientific.
- Bollen, K.A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1), 605–634.
- Bollen, K.A., & Medrano, J.D. (1998). Who are the Spaniards? Nationalism and identification in Spain. *Social Forces*, 77(2), 587–621.
- Bolton, R., & Drew, J. (1991). A longitudinal analysis of the impact of service changes on customer attitudes. *Journal of Marketing*, 55(1), 1–9.
- Brady, M.K., & Cronin, J.J., Jr. (2001). Some new thoughts on conceptualizing perceived service quality: A hierarchical approach. *Journal of Marketing*, 65(3), 34–49.
- Chen, C.T. (2000). Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets and Systems*, 114(1), 1–9.
- Cox, E.P. (1980). The optimal number of response alternatives for a scale: a review. *JMR, Journal of Marketing Research*, 17(4), 407–422.
- Crompton, J.L., & MacKay, K.L. (1989). User's perceptions of the relative importance of service quality dimensions in selected public recreation programs. *Leisure Sciences*, 11(4), 367–375.
- Crompton, J.L., MacKay, K.J., & Fesenmaier, D.R. (1991). Identifying dimensions of service quality in public recreation. *Journal of Park and Recreation Administration*, 9(3), 15–27.
- Cronin, J.J., Jr., & Taylor, S.A. (1992). Measuring service quality: a reexamination and extension. *Journal of Marketing*, 56(3), 55–68.
- Dabholkar, P.A., Thorpe, D.I., & Rentz, J.O. (1996). A measure of service quality for retail stores: Scale development and validation. *Journal of the Academy of Marketing Science*, 24(1), 3–16.
- Drolet, A., & Morrison, D.G. (2001). Do we really need multiple-item measures in service research? *Journal of Service Research*, 3(3), 196–204.
- Ekinci, Y. (2001). The validation of the generic service quality dimension: An alternative approach. *Journal of Retailing and Consumer Services* 8(6), 311–324.
- Ferrando, P.J. (2003). A Kernel density analysis of continuous typical-response scales. *Educational and Psychological Measurement*, 63(5), 809–824.
- Getz, D., O'Neill, M., & Carlsen, J. (2001). Service quality evaluation at events through service mapping. *Journal of Travel Research*, 39(2), 380–390.
- Gil-Aluja, J. (1999). *Investment in uncertainty*. London: Kluwer Academic Publishers.
- Grönroos, C. (1982). *Strategic management and marketing in the service sector*. Helsinki: Swedish School of Economics and Business Administration.
- Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing*, 18(4), 36–44.
- Guha, D. Chakraborty, D. (2010). A new approach to fuzzy distance measure and similarity measure between two generalized fuzzy numbers. *Applied Soft Computing Journal*, 10(1), 90–99.
- Hayduk, L.A., & Glaser, D.N. (2000). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling*, 7(1), 1–35.
- Hayduk, L.A., Pazderka-Robinson, H., Cummings, G., Boadu, K., Verbeek, E., & Perks, T. (2007). The weird world and equally weird measurement models: Reactive indicators and the validity revolution. *Structural Equation Modeling*, 14(2), 280–310.
- Hernández, A. (2001). Un cuestionario para evaluar la calidad en programas de actividad física [A questionnaire to assess the quality of physical activity programs]. *Revista de Psicología del Deporte*, 10(2), 179–196.
- Herzberg, F., Mausner, B., & Snyderman, B.B. (1959). *The Motivation to work*. New York: John Wiley.
- Hofmans, J., Theuns, P., & Mairesse, O. (2007). On the impact of the number of response categories on linearity and sensitivity of 'Self Anchoring Scales'. A Functional Measurement approach. *Methodology*, 3(4), 160–169. [http://ppw.kuleuven.be/okp/people/Jori\\_Hofmans/](http://ppw.kuleuven.be/okp/people/Jori_Hofmans/)

- Hofmans, J., Theuns, P., & Van Acker, F. (2009). Combining quality and quantity. A psychometric evaluation of the self-anchoring scale. *Quality & Quantity*, 43(5), 703-716.
- Howat, G., Absher, J., Crilley, G., & Milne, I. (1996). Measuring customer service quality in sports and leisure centers. *Managing Leisure*, 1(2), 77-89.
- Huang, C.C., Chu, P.Y., & Chiang, Y.H. (2008). A fuzzy AHP application in government-sponsored R & D project selection. *Omega*, 36(6), 1038-1052.
- Javaras, K.N., & Ripley, B.D. (2007). An "unfolding" latent variable model for Likert attitude data: drawing inferences adjusted for response style. *Journal of the American Statistical Association*, 102(478), 454-463.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.
- Karatepe, O.M., Yavas, U., & Babakus, E. (2005). Measuring service quality of banks: Scale development and validation. *Journal of Retailing and Consumer Services*, 12(5), 373-383.
- Kilpatrick, F.P., & Cantril, H. (1960). Self-anchoring scaling: A measure of individuals' unique reality worlds. *Journal of Individual Psychology*, 16, 158-173.
- Kelly, S.W., & Turley, L.W. (2001). Consumer perceptions of service quality attributes at sporting events. *Journal of Business Research*, 54(2), 161-166.
- Kim, D., & Kim, S.Y. (1995). QUESC: an instrument for assessing the service quality of sport centers in Korea. *Journal of Sport Management*, 9(2), 208-220.
- Ko, Y.J., Durrant, S.M., & Mangiantini, J. (2008). Assessment of services provided to NCAA Division I athletes: Development of a model and instrument. *Sport Management Review*, 11(2), 193-214.
- Ko, Y.J., & Pastore, D.L. (2004). Current issues and conceptualizations of service in the recreational sport industry. *Sport Marketing Quarterly*, 13(3), 159-167.
- Ko, Y.J., & Pastore, D.L. (2005). A hierarchical model of service quality for the recreational sport industry. *Sport Marketing Quarterly*, 14(2), 84-97.
- Levy, P.S., & Lemeshow, S. (1999). *Sampling of populations: Methods and applications* (3rd ed.). New York: John Wiley & Sons, Inc.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives de Psychologie*, 140, 1-55.
- Lissitz, R.W., & Green, S.B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *The Journal of Applied Psychology*, 60(1), 10-13.
- Liu, P., & Wang, T. (2008). Research on risk evaluation in supply chain based on grey relational method. *Journal of Computers*, 3(10), 28-35.
- Lozano, M.C., & Fuentes, F. (2003). *Tratamiento borroso del intangible en la valoración de empresas de Internet*. Retrieved from <http://www.eumed.net/cursecon/libreria/pedir-clg-ffm.html>
- MacKay, K.J., & Crompton, J.L. (1988). A conceptual model of consumer evaluation of recreation service quality. *Leisure Studies*, 7(1), 41-49.
- Martínez, L., & Martínez, J.A. (2007). Measuring perceived service quality in urgent transport service. *Journal of Retailing and Consumer Services*, 14(1), 60-72.
- Martínez, J.A., & Martínez, L. (2010). Some insights on conceptualizing and measuring service quality. *Journal of Retailing and Consumer Services*, 17(1), 29-42.
- Martínez, L., & Martínez, J.A. (2009). Does ISO 9000 Certification affect consumer perceptions of the service provider? *Managing Service Quality*, 19(2), 140-161.
- Martínez, J.A., & Ruiz, M. (in press). D-test: A new test for analysing scale invariance using symbolic dynamics and symbolic entropy. *Methodology*.
- McDonald, M.A., Sutton, W.A., & Milne, G.R. (1995). TEAMQUAL: Measuring service quality in professional sports. *Sport Marketing Quarterly*, 4(2), 9-15.
- Milne, G.R., & McDonald, M.A. (1999). *Sport marketing: Managing the exchange process*. Sudbury, MA: Jones and Bartlett.
- Morales, V., Hernández, A., & Blanco, A. (2005). Evaluación de la calidad en los programas de actividad física [Quality evaluation in physical activity programs]. *Psicothema*, 17(2), 311-317.



- Nejati, M., Nejati, M., & Shafaei, A. (2009). Ranking airlines' service quality factors using a fuzzy approach: study of the Iranian society. *International Journal of Quality & Reliability Management*, 26(3), 247–260.
- Papadimitriou, D.A., & Karateroliotis, K. (2000). The service quality expectations in private sport and fitness centers: A reexamination of the factor structure. *Sport Marketing Quarterly*, 9(3), 157–164.
- Parasuraman, A., Zeithaml, V.A., & Berry, L.L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(4), 41–50.
- Parasuraman, A., Zeithaml, V.A., & Berry, L.L. (1988). SERVQUAL: A multiple item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12–40.
- Parasuraman, A., Zeithaml, V.A., & Berry, L.L. (1994). Alternative scales for measuring service quality: A comparative assessment based on psychometric and diagnostic criteria. *Journal of Retailing*, 70(3), 193–194.
- Preston, C.C., & Colman, A.M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1–15.
- Rajasekharan, S., & Vijayalakshmi, G.A. (2003). *Neural networks, Fuzzy logic and genetic algorithms*. New Delhi: Prentice Hall.
- Rossiter, J.R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 1–31.
- Rust, R.T., & Oliver, R.L. (1994). Service quality: Insights and managerial implications from the frontier. In R.T. Rust & R.L. Oliver (Eds.), *Service quality: New directions in theory and practice* (pp. 1–19). London: Sage Publications.
- Saris, W.E., & Gallhofer, I. (2007). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods*, 1(1), 29–43.
- Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.
- Shonk, D.J., & Chelladurai, P. (2008). Service quality, satisfaction, and intent to Return in event sport tourism. *Journal of Sport Management*, 22(5), 587–602.
- Taylor, S.A., Sharland, A., Cronin, J.J., Jr., & Ballard, W. (1993). Recreational service quality in the international setting. *International Journal of Service Industry Management*, 4(4), 68–86.
- Theodorakis, N., Kambitsis, C., Laios, A., & Koustelios, A. (2001). Relationship between measures of service quality and satisfaction of spectators in professional sport. *Managing Service Quality*, 11(6), 431–438.
- Tsai, M.T., Wu, H.L., & Liang, W.K. (2008). Fuzzy Decision Making for Market Positioning and Developing Strategy for Improving Service Quality in Department Stores. *Quality & Quantity*, 42(3), 303–319.
- Tsitskari, E., Tsiotras, D., & Tsiotras, G. (2006). Measuring service quality in sport services. *Total Quality Management & Business Excellence*, 17(5), 623–631.
- Tseng, M.L. (2009). Using the extension of DEMATEL to integrate hotel service quality perceptions into a cause–effect model in uncertainty. *Expert Systems with Applications*, 36, 9015–9023.
- Wakefield, K.L., & Blodgett, J.G. (1994). The importance of Servicescapes in leisure service setting. *Journal of Services Marketing*, 8(3), 66–76.
- Wakefield, K.L., Blodgett, J.G., & Sloan, H.J. (1996). Measurement and management of the Sportscape. *Journal of Sport Management*, 10(1), 15–31.
- Wallsten, T.S., Budescu, D.V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, 39(2), 176–190.
- Weinstein, N.D., & Diefenbach, M.A. (1997). Percentage and verbal-category scales for assessing personal risk likelihood. *Health Education Research*, 12(1), 139–141.
- Weng, L.J., & Cheng, C.P. (2000). Effects of response order on Likert-type scales. *Educational and Psychological Measurement*, 60(6), 908–924.
- Windschitl, P.D., & Wells, G.L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, 2(4), 343–364.

- Wright, B.A., Duray, N., & Goodale, T.L. (1992). Assessing perceptions of recreation center service quality. An application of recent advancements in service quality research. *Journal of Park and Recreation Administration*, 10(3), 33–47.
- Yim, C. K. (Bennett), Tse, D. K., & Chan, K. W. (2008). Strengthening customer loyalty through intimacy and passion: Roles of customer-firm affection and customer-staff relations in services. *Journal of Marketing Research*, 45(6), 741–756.
- Zadeh L. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.
- Zeithaml, V.A. (1988). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *Journal of Marketing*, 52(3), 2–22.
- Zwick, R., Carlstein, E., & Budescu, D.V. (1987). Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning*, 1, 221–242.
- Zwick, R., & Wallsten, T.S. (1989). Combining stochastic uncertainty and linguistic inexactness: Theory and experimental evaluation of four fuzzy probability models. *International Journal of Man-Machine Studies*, 30(1), 69–111.

## Appendix

The stability of fuzzy numbers is an important concern, because low sample sizes cannot be enough to form an adequate fuzzy number. The form of a triangular fuzzy number depends on three parameters  $\tilde{A}_i = \{a'_i, b'_i, c'_i\}$  which they are the vertexes of the triangle. As we have explained,  $b'_i = \sum p_{ij}x_{ij}$  where  $p_j$  is the probability of an individual  $j$  responding the value  $x_j^i$  for each linguistic term  $i \in m$ . Finally,  $a'_i = \min(x_{ij})$  &  $c'_i = \max(x_{ij})$ . Therefore,  $b'_i$  is the expected value of the distribution. If we assume that  $\hat{b}_i$  is the estimate of  $b'_i$ , then  $\hat{b}_i$  is a sample statistic, and a confidence interval can be computed (Levy & Lemeshow, 1999), using the following expression (without counting with the finite population factor):

$$\hat{b}_i - t_{\alpha,v} \frac{S}{\sqrt{n}} \leq b'_i \leq \hat{b}_i + t_{\alpha,v} \frac{S}{\sqrt{n}} = \hat{b}_i - \text{Error} \leq b'_i \leq \hat{b}_i + \text{Error}$$

Consequently, the size of the error highly depends on the sample size. As the measure's units are normalized to the unit interval, then we can fix a maximum admissible error. A reasonable size for that error would be the 5% of the length of the scale (i.e., 0.05 units).

Therefore we may calculate the size of the error for each  $\hat{b}_i$ . As we have stated, only the term “good” has a reasonable sample size to form a consistent fuzzy number. Therefore, for the first sample:  $\text{Error} = 0.033$ , and for the second sample:  $\text{Error} = 0.028$ . If we pool both standard deviations to assume an average variance for all the possible linguistic terms, then we may compute the sample size necessary to estimate  $\hat{b}_i$  within the admissible error. Under these assumptions, we need 22 cases per linguistic term to fulfill the requirement.

None of the remaining linguistic terms of this research meet with the criteria of the minimum sample size.