The 5-Item Likert Scale and Percentage Scale Correspondence with Implications for the Use of Models with (Fuzzy) Linguistic Variables

MAZUREK, JIŘÍ

School of Business Administration in Karvina
Silesian University in Opava (the Czech Republic)
Correo electrónico: mazurek@opf.slu.cz

PÉREZ RICO, CRISTINA

Facultad de Ciencias Administrativas, DESODEH
Escuela Politécnica Nacional (Ecuador)
Correo electrónico: cristina.perez@epn.edu.ec

FERNÁNDEZ, CARLOS

Escuela de Negocios. Universidad de las Américas (Ecuador)
Correo electrónico: carlos.fernandez.garcia@udla.edu.ec

MAGNOT, JEAN-PIERRE

LAREMA, University of Angers (France)
Correo electrónico: magnot@math.univ-angers.fr

MAGNOT, TRISTAN

Institut de Management Public et de Gouvernance Territoriale (France)
Correo electrónico: tristan.magnot@etu.univ-amu.fr

ABSTRACT

The aim of this paper is to examine how people perceive correspondence between the 5-item Likert scale and the percentage scale (the LS-PS correspondence thereinafter). Are all five items of the Likert scale equidistant? Do people use the same scale when evaluating different objects? Are men and women different? Are people from different countries / cultures different? The method of the study was a questionnaire with 661 participating respondents altogether from the Czech Republic, Ecuador, and France. The results indicate that the 5-item Likert scale is neither equidistant, nor symmetrical. Furthermore, there are (highly) statistically significant differences in the LS-PS correspondence with respect to location, age, or gender of respondents. The results can be used as an input for more precise decision-making modeling associated with (fuzzy) linguistic variables.
Keywords: decision-making, evaluation, fuzzy linguistic variables, international study, Likert scale.
JEL classification: D12; D91.
MSC2010: 62P25; 68P99; 90B50.

La escala de Likert de 5 ítems y la correspondencia de la escala de porcentaje con implicaciones para el uso de modelos con variables lingüísticas (difusas)

RESUMEN

El objetivo de este trabajo es examinar cómo las personas perciben la correspondencia entre la escala Likert de 5 ítems y la escala de porcentaje (la correspondencia LS-PS en adelante). ¿Los cinco elementos de la escala Likert son equidistantes? ¿La gente usa la misma escala al evaluar diferentes objetos? ¿Se diferencian hombres y mujeres? ¿Son diferentes las personas de diferentes países / culturas? El método del estudio fue un cuestionario con 661 encuestados participantes de la República Checa, Ecuador y Francia. Los resultados indican que la escala Likert de 5 ítems no es ni equidistante ni simétrica. Además, existen diferencias (altamente) estadísticamente significativas en la correspondencia LS-PS con respecto a la ubicación, edad o género de los encuestados. Los resultados se pueden utilizar como entrada para un modelo de toma de decisiones más preciso asociado con variables lingüísticas (difusas).

1. Introduction.

Evaluation of objects such as goods, services, or companies on various scales is almost omnipresent. However, it is difficult to measure human attitude, character and personality because of its subjective nature (Prasad, 2016), and to transform subjective opinions into objective reality (Joshi, Kale, Chandel & Pal, 2015). The use of scales enables comparing, rating, or ranking objects, which is important in many areas of business, science, entertainment, or personal life. The scales used for evaluation are usually numerical or linguistic. In the former case, objects are usually evaluated on scales from 1 to 5 (typically in the form of 1 star to 5 stars), from 1 to 10, or in percents from 0 to 100%. This approach is followed by many Internet platforms such as Booking.com, which assigns each accommodation option values from 1 to 10; Goodreads.com, where its users rate books from 1 star to 5 stars; IMDb.com, where a movie fan can find rating of his/her favorite movie from 1 to 10, etc.

In the latter case, objects are assigned "labels", such as from A to F when evaluating students' performance in the USA and majority of EU countries, from A to H for energy consumption of houses, credit ratings from A to C or D, and so on. In the case of the linguistic scales, its values (linguistic terms) are called categories, and it is required that the set of all categories of a given scale is ordinal, which means categories are ordered from the worst (least desirable) ones to the best (most desirable) ones.

A Likert scale, introduced by Rensis Likert (Likert, 1932), is a bipolar scaling method, which measures positive or negative response to a given statement. The scale is very popular among researchers in psychology, sociology, pedagogy, marketing, or business that involves questionnaires. A Likert scale is usually treated as ordinal or interval data. It is usually constructed from an odd number (3, 5, 7, or 9) of categories, where the medium category expresses ambivalence, indecisiveness, or lack of an opinion of a respondent, while the first and the last category express an extreme opinion. Typically, the five-item Likert scale has the following form: {very good; good; average; poor; very poor}, or {strongly agree; agree; neither agree, nor disagree; disagree; strongly disagree}. It has become common practice to treat this scale as a rating scale and assume that the equal intervals hold between the response categories, see e.g. Blaikie (2003). However, as Creswell (2008) points out, we have no guarantee that intervals are equal indeed. The assumption that the "distance" between each successive pair of categories is the same, is going to be challenged in this paper. More on Likert scales, their proper use, statistical properties, or applications can be found in Allen and Seaman (2007), Blaikie (2003), Clason and Dormody (1994), Creswell (2008), Dawes (2008), Edwards and Kenney (1946), Hasson and Arnetz (2005), Joshi et al. (2015), Norman (2010), Traylor (1983), Prasad (2016), or Willits et al. (2016).

Finally, Booking.com is the example of a platform that combines the numerical and linguistic approach. According to Booking.com, an accommodation rated 9+ is "superb", 7-9 means "good", "okay" is 5-7, "poor" 3-5, and "very poor" 1-3. The origin of this correspondence remains unclear, since there is no study linking linguistic terms such as "superb" with the percentage scale.

Apart from Internet platforms, linguistic variables are extensively used in decision-making modeling under uncertainty. Usually, the models express linguistic terms via triangular (or other) fuzzy numbers, and Likert scales are assumed to be equidistant, see e.g. Carrasco et al. (2012), Casola et al. (2005), Holeček and Talašová (2010), Lima et al. (2016), Lin and Yeh (2012), or Yan et al. (2017). However, if individuals do not perceive Likert scales to be equidistant, this input assumption might lead to unwanted distortion of results.

Therefore, the aim of this paper is to examine a Likert scale and percentage scale correspondence (the LS-PS correspondence in short) to find answers to the following questions: is 5-item Likert scale perceived to be equidistant by respondents (decision makers)? Are there any differences with respect to location, gender, or age of respondents? Do respondents use the same (universal) scale for evaluation? To find answers to the aforementioned questions, several research hypotheses were formulated and statistically tested. The method of the study was an experiment carried out via a questionnaire.

The presented study is the first of its kind, but its results are not only of theoretical worth. On the contrary, it has important practical implications. The results can be used as an input to decision-making models with (fuzzy) linguistic variables, improving their accuracy.

The paper is organized as follows. In Section 2 the research problem is presented. In Section 3 data and the method of the study are described. Section 4 provides results of the study and in Section 5 a more accurate 5-item Likert scale for modeling with triangular fuzzy numbers is proposed. Conclusions close the article.

2. The problem and research hypotheses.

To date, there is no empirical study known to the authors on how exactly people perceive linguistic scales as well as their identification with the percentage scale. Moreover, there is no evidence on whether people use one (general) scale for evaluation or different scales in different situations. Likert scales are assumed to be symmetrical and equidistant. But how much is such an assumption realistic? Do all people use the same scale? Do they use such a scale in different situations and for different objects? Is there no difference between them for respondents with respect to country/culture, gender, or age?

To find out the answers to the aforementioned questions, the relationship between 5-item Likert scale and the percentage scale was examined by the experiment described in the next section so that the following null hypotheses can be statistically tested:

- H_{01}: Items on the 5-item Likert scale are evenly distributed (are equidistant) with respect to the percentage scale.

- H_{02}: Individuals from different countries use the same percentage scale for the corresponding 5-item Likert scale (LS-PS correspondence).

- H_{03}: Men and women use the same LS-PS correspondence.

- H_{04}: Younger and older respondents use the same LS-PS correspondence.

Further, one research question (Q1) was examined:

Q1: Do individuals use one (general, universal) scale for the evaluation of different objects (goods or services)?

3. Data and method.

The experiment linking the 5-item Likert scale with the percentage scale was carried out during 2017-2018. Table 1 summarizes the locations and the number of respondents.

The method of a study was a questionnaire disseminated online in Spanish, French, and Czech languages, respectively. High school students from Clermont-Ferrand and university students from the locations such as Karvina, Quito, Cuenca, Marseille, and Angers (see Table 1) formed the set of 661 respondents. The respondents were instructed to assign each of 5 categories of the Likert scale (*very good – good – average – poor – very poor*) an interval in percents corresponding to their perception of what is "very good", "good" and so on, when evaluating two items: a movie and a cell phone. Thus, each student provided two LS-PS correspondences, which amounts to 1,322 LS-PS correspondences altogether. Naturally, the intervals assigned to the five items had to cover the entire interval [0, 100%] without overlapping:

$$\cup I_i = [0,100\%] \wedge (I_i \cap I_j) = \varnothing; \forall i,j \in \{1,2,...,5\}, i \neq j. \tag{1}$$

Let β_1, β_2, β_3, and β_4 denote the boundary values between "very poor" and "poor", "poor" and "average", "average" and "good" and finally "good" and "very good" categories, respectively. Further, let $\beta_0 = 0$ and $\beta_5 = 100$.

Clearly, we have:

$$\beta_i > \beta_j \text{ for } i > j. \tag{2}$$

Let δ_1, δ_2, δ_3, δ_4, and δ_5 correspond to the "lenghts" of intervals corresponding to "very poor", "poor", "average", "good", and "very good" categories, respectively.

Obviously, we have:

$$\delta_i = \beta_i - \beta_{i-1} \quad , i = 1,...,5. \tag{3}$$

The questionnaires that didn't satisfy condition (1) or (2) were discarded (less than 2% altogether).

To test the null hypotheses H_{02}-H_{04}, the independent two-sample t-tests with unequal variances (with normality tested as well) and one-factor ANOVA were applied (see the next section). Hypothesis H_{01} was tested via a paired two-sample t-test and one-factor ANOVA. Testing was performed via MS Excel and Gretl.

Table 1. The summary of respondents' nationality, location, and gender.

Country	University/City	Number of respondents (men/women)
Czech Rep.	School of Business Administration/ Karvina	210 (70/140)
Ecuador	Faculty of Administrative Sciences, Escuela Politécnica Nacional/ Quito	300 (94/206)
	Faculty of Economics and Administrative Sciences, Universidad de Cuenca; Universidad Politénica Salesiana / Cuenca	
France	University of Angers; Lycée Jeanne d'Arc Clermont Ferrand; AIX-Marseille Université	151 (60/91)

Source: Own elaboration.

4. Results.

The results of the study are summarized in Tables 2-6, where the overall results and results with respect to geographic location, gender, and age of respondents are provided.

As can be seen from Tables 2 and 3, respondents didn't divide, on average, the 5-item Likert scale uniformly with boundaries $\beta_i = \{0,20,40,60,80,100\}$ and width $\delta_i = 20$ as one might expect.

Another interesting feature were regional differences in β_i (see Table 2). The β_i values of respondents from Ecuador were the highest, meaning the respondents were the most "demanding" in their evaluations. On the other hand, the values from respondents in France were the lowest. Also, respondents from France assigned the largest width (δ_5) to the category "very good", while respondents from the Czech Republic and Ecuador assigned the largest width (δ_1) to the category "very poor". The differences between France and the other two countries could be caused by more heterogenous sample of French respondents with respect to age and education.

Table 2. Average upper boundaries (β_i) of the 5-item Likert scale (with standard deviation in brackets) for all locations.

Linguistic term/upper boundaries	All countries (CZE+ECU+FRA)	CZE	ECU	FRA
Very good	100 (0)	100 (0)	100 (0)	100 (0)
Good (β_4)	81.43 (9.78)	82.84 (7.76)	83.07 (8.09)	76.20 (13.01)
Average (β_3)	61.64 (11.42)	63.62 (9.53)	63.76 (11.17)	54.67 (11.60)
Poor (β_2)	41.47 (11.58)	43.98 (9.25)	43.48 (11.37)	33.98 (11.78)
Very poor (β_1)	22.60 (10.56)	24.84 (10.28)	23.62 (10.45)	17.44 (9.45)

Source: Own elaboration.

Table 3. Average width (δ_i) of items of the 5-item Likert scale (with standard deviation in brackets) for all locations.

Linguistic term/upper boundaries	All countries (CZE+ECU+FRA)	CZE	ECU	FRA
Very good (δ_5)	18.54 (9.78)	17.16 (8.01)	16.94 (8.09)	23.80 (13.01)
Good (δ_4)	19.79 (7.51)	19.22 (6.90)	19.31 (6.79)	21.52 (9.22)
Average (δ_3)	20.17 (7.04)	19.64 (6.92)	20.28 (6.72)	20.69 (7.76)
Poor (δ_2)	18.87 (6.91)	19.15 (6.47)	19.86 (6.61)	16.54 (7.51)
Very poor (δ_1)	22.60 (10.56)	24.84 (10.28)	23.62 (10.45)	17.44 (9.45)

Source: Own elaboration.

As for gender diferences (Table 4), men assigned higher values to all β_i than women on altogether, but also for each country separately.

Table 4. Average upper boundaries (β_i) of Likert scales (with standard deviation in brackets) for all locations with respect to gender.

Linguistic term/upper boundaries	All countries Women only	All countries Men only	CZE Women only	CZE Men only	ECU Women only	ECU Men only	FRA Women only	FRA Men only
Very good	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)
Good (β_4)	80.88 (10.27)	82.23 (9.58)	82.85 (8.1)	84.04 (7.0)	82.95 (8.38)	83.33 (7.43)	74.08 (13.74)	77.49 (15.12)
Average (β_3)	60.94 (11.87)	62.82 (10.74)	62.81 (9.8)	65.24 (8.9)	63.76 (11.59)	63.78 (10.18)	51.70 (10.83)	58.07 (12.64)
Poor (β_2)	40.74 (11.70)	42.80 (11.28)	43.15 (9.0)	45.65 (9.6)	43.42 (11.67)	43.62 (10.65)	30.95 (10.17)	37.98 (12.75)
Very poor (β_1)	22.39 (10.37)	23.07 (10.94)	24.18 (9.9)	26.15 (10.9)	23.97 (10.56)	22.85 (10.14)	16.06 (7.93)	20.04 (11.24)

Source: Own elaboration.

Further, only 23.3% individuals used the same scale in the evaluation of a mobile phone and a movie (Table 5).

Table 5. Did respondents use the identical correspondence between Likert and percentage scales for a movie and a cell phone?

Identical correspondence (%)	All countries	CZE	ECU	FRA
Yes	23.27	15.24	21.94	37.09
No	76.73	84.76	78.06	62.91

Source: Own elaboration.

Finally yet importantly, younger respondents provided lower values of upper boundaries β_i than their older counterparts (Table 6).

Table 6. Younger respondents versus older respondents.

Linguistic term/upper boundaries	Younger respondents (15-29 years)	Older respondents (30 and more)
β_1	16.26 (8.64)	19.35 (10.41)
β_2	31.38 (10.36)	36.66 (12.74)
β_3	51.82 (11.55)	57.21 (11.90)
β_4	73.58 (13.64)	77.69 (14.98)

Source: Own elaboration.

4.1. Equidistance of the 5-item Likert scale.

The first problem examined in the presented study was the problem of (alleged) equidistance of the 5-item Likert scale. The null hypothesis states all five categories have the same "width".

- The null hypothesis H_{01}: $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5$.

The hypothesis was statistically tested via a paired two-sample t-test with unequal variances (in MS Excel, after checking for data normality), where the test of all 10 pairs of δ_i and δ_j, $i \neq j$, $i, j \in \{1,2,3,4\}$ was performed with its results being shown in Table 7a.

Table 7a. The t-test statistics (p values) for the hypothesis H_{01}.

p values	δ_1	δ_2	δ_3	δ_4	δ_5
δ_1	-	$<10^{-21}$	$2 \cdot 10^{-9}$	$2.34 \cdot 10^{-11}$	$3 \cdot 10^{-17}$
δ_2		-	$2.25 \cdot 10^{-6}$	0.0024	0.440
δ_3			-	0.205	$1.4 \cdot 10^{-5}$
δ_4				-	0.0008
δ_5					-

Source: Own elaboration.

Based on p-values from Table 7a, the null hypothesis H_{01} was rejected. Thus, 5-item Likert scale was not percieved equidistant by respondents. Also, we applied ANOVA to test the hypothesis H_{01} (Table 7b) with the same result: H_{01} was rejected.

Table 7b. ANOVA test for the H_{01} hypothesis.

Variability source	SS	Diff.	MS	F	P	F krit
Among samples	10752.92	2	5376.462	51.91338	0	3.002547
All samples	136603.6	1319	103.566			
Total	147356.5	1321				

Source: Own elaboration.

If the relation $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5$ held, categories "very good", "good", etc., would have the same width $\delta = 20$. We tested that assumption by a one-sample z-test. The p-value $p < 0.01$ was found only for δ_1.

4.2. National (geographic, cultural) differences in LS-PS correspondence.

Since respondents originated from different countries and cultures, the following null hypothesis was tested:

- H_{02}: Individuals from different countries use the same percentage scale

  ($\beta_i^{ECU} = \beta_i^{CZE} = \beta_i^{FRA}; \forall i$) for the corresponding 5-item Likert scale (LS-PS correspondence).

The hypothesis was tested (for each β_i separately) by one-factor ANOVA, where the independent factor was a country of respondents: Czech Republic, Ecuador, and France. The results of ANOVA (performed in MS Excel) are provided in Table 8.

The null hypothesis H_{02} was rejected at least at $p = 10^{-34}$ (!) level (other p-values were even smaller and rounded by MS Excel to 0).

Table 8. Results of ANOVA for all β_i.

Boundary	F-stat	p
β_1	51.91	10^{-34}
β_2	93.56	0
β_3	81.69	0
β_4	61.16	0

Source: Own elaboration.

4.3. Gender differences in LS-PS correspondence.

To test (possible) differences between men and women with respect to LS-PS correspondence, the following null hypothesis was formulated:

- H_{03}: Men and women use the same LS-PS correspondence ($\beta_i^{men} = \beta_i^{women}, \forall i$).

For the test, a two-sample unpaired t-test with unequal variances was performed (for each β_i separately), and the results are shown in Table 9a. Since $\beta_i^{men} \neq \beta_i^{women}$ for $i = 2, 3$ and 4, the null hypothesis H_{03} was rejected at $p = 0.0018$ level.

Table 9a. Results of t-tests for H_{03} hypothesis.

Boundary	Mean Men/Women	Variance Men/Women	Observ. Men/Women	t-stat.	p
β_4	82.227/80.876	91.693/105.371	448/874	2.3764	0.01768
β_3	62.819/60.44	115.280/140.802	448/874	2.9060	0.00374
β_2	42.803/40.736	127.294/136.820	448/874	3.1232	0.00184
β_1	23.071/22.388	119.698/107.500	448/874	1.0966	0.2731

Source: Own elaboration.

In addition, gender differences in all three countries (separately) were examined. The following Table 9b shows p-values of the null hypothesis H_{03} tested by two-sample t-test.

At $p = 0.01$ level, the null hypothesis H_{03} was rejected in the case of French respondents.

Table 9b. The p-values for the hypothesis H_{03}.

p	β_1	β_2	β_3	β_4
CZE	0.073	0.011	0.01	0.019
ECU	0.217	0.837	0.979	0.573
FRA	0.001	$< 10^{-6}$	$< 10^{-5}$	0.006

<div align="center">Source: Own elaboration.</div>

4.4. Age differences in LS-PS correspondence.

To test possible differences between younger and older respondents, respondents were divided into two age groups: 15-29, and 30 and more.

- H_{04}: Younger and older respondents use the same LS-PS correspondence ($\beta_i^{younger} = \beta_i^{older}$, $\forall i$).

Since the data from France were the only one including older respondents (above 29 years of age), they were the only data applicable for the testing of hypothesis H_{04}. The result of two-sample unpaired t-test with unequal variances (for each β_i separately) is shown in Table 10.

The null hypothesis H_{04} was rejected at $p = 0.01$ level.

Table 10. Results of the t-tests for H_{04} hypothesis.

Boundary	Mean Younger/Older	Variance Younger/Older	Observ. Younger/Older	t-stat.	p
β_1	16.259/19.350	75.08/109.18	166/136	2.7846	0.00574
β_2	31.379/36.664	107.91/163.49	166/136	3.9196	0.00011
β_3	51.819/57.207	134.14/142.67	166/136	3.9859	0.00009
β_4	73.578/77.693	187.13/226.16	166/136	2.4844	0.01355

<div align="center">Source: Own elaboration.</div>

4.5. Uniqueness of the 5-item Likert scale.

Many models based on (fuzzy) linguistic variables assume the Likert scale to be equidistant, which means the scale is also invariant with respect to a subject of evaluation. Consequently, humans are supposed to use one (universal) scale for evaluation. Therefore, the following question was examined:

- Q$_1$: Do individuals use one (general, universal) scale for the evaluation of different objects (goods, services, etc.)?

According to the empirical data acquired in this study, only 23.27% of respondents used the same correspondence between percentage scale and Likert scale for a movie and a cell phone. Thus, the meaning of linguistic terms "very good", "good", etc., is not universal and varies with respect to a subject of evaluation.

The answer to the question Q$_1$ is therefore negative.

5. The 5-item Likert scale and its representation by triangular fuzzy numbers.

In the context of a decision-making, linguistic variables are used to model uncertainty present in many real-world problems. As such, they are often expressed in terms of fuzzy sets. Here, we recall two useful definitions:

Definition 1 (the fuzzy set). Let U denote the universal set of discourse. Then, a fuzzy set $A = \{(x, \mu(x)); x \in U\}$ on U is defined in terms of the membership function μ: $\mu_A(x): U \to [0,1]$.

Definition 2 (the triangular fuzzy number). Let $a, b, c \in R; a < b < c$. Then, the triangular fuzzy number (*TFN*) is a triplet (*a*, *b*, *c*) with the membership function given as follows:

$$\mu = \begin{cases} \dfrac{x-a}{b-a}, x \in [a,b], \\ \dfrac{c-x}{c-b}, x \in [b,c], \\ 0, x < a \vee x > c. \end{cases} \tag{4}$$

The use of Likert scales in the context of fuzzy sets was introduced in Hesketh et al. (1988), since fuzzy approach can provide more realistic framework for decision making with ordinal (linguistic) variables. Likert scales are usually represented by triangular fuzzy numbers as can be seen in Figures 1 and 2. Perhaps the most striking feature of both pictures is items' symmetry and equidistance. There are hundreds of similar pictures in the literature assuming Likert scales to be symmetrical and equidistant.

The equidistant 5-item Likert scale modeled with triangular fuzzy numbers (4) in the symmetrical form used in Carrrasco et al. (2012), Holeček and Talašová (2010), or Yan et al. (2017), takes the form shown in Table 11 (column 2), where each term of the Likert scale is represented by one TFN.

Figure 1. Symmetrical 5-item fuzzy Likert scale.
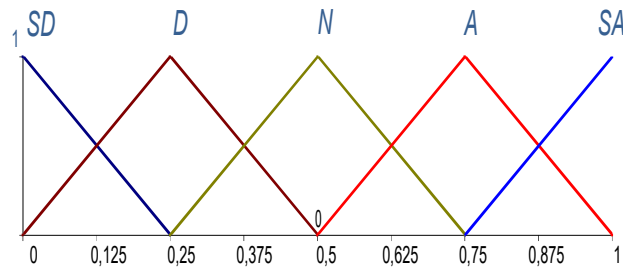
Source: Carrasco et al. (2012).

Figure 2. Symmetrical 7-item fuzzy Likert scale.
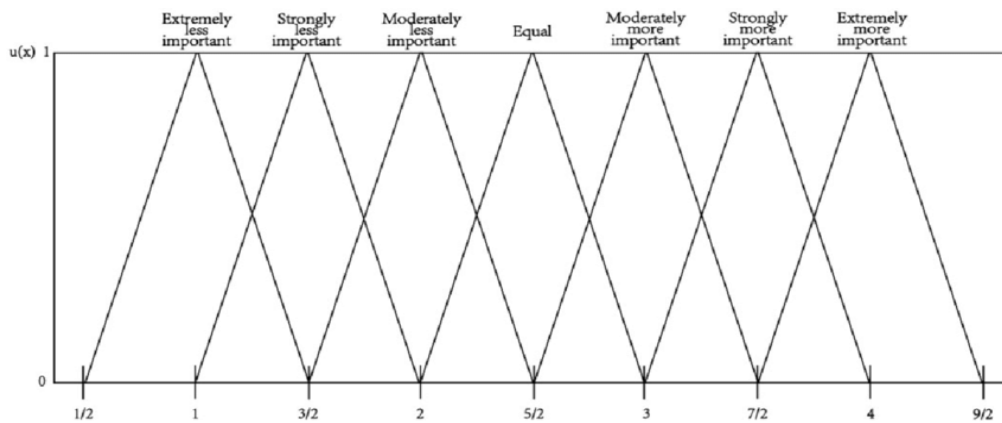
Source: Yan et al. (2017).

However, according to our findings, 5-item Likert scale is not equidistant, nor symmetrical. A new, more realistic expression of the 5-item Likert with TFNs, based on the empirical data acquired by the presented study, is provided in Table 11, the 3rd column. Although the numerical differences between both sets of TFNs are not large, they may contribute to more accurate modeling and decision making involving 5-item Likert scales.

Table 11. Usual equidistant 5-item Likert scale expressed by TFNs (2nd column), and corrected TFNs (3rd column).

Scale item	TFNs for equidistant scale	Corrected TFNs
Very poor	$(a, b, c) = (0;0;20)$	$(a, b, c) = (0;0;22.6)$
poor	$(0;20;40)$	$(0;22.6;41.5)$
average	$(20;40;60)$	$(22.6;41.5;61.6)$
good	$(60;80;100)$	$(61.6;81.4;100)$
Very good	$(80;100;100)$	$(81.4;100;100)$

Source: Own elaboration.

Of course, triplets (a, b, c) in TFNs' definition can also be further adjusted for gender, age, or geographic location since differences in the perception of the LS-PS correspondence with respect to the aforementioned factors were found statistically (highly) significant. These refinements may lead to even more precise results.

6. Conclusions.

The aim of the paper was to examine empirical relationship between the 5-item Likert scale and the percentage scale. The results indicate that decision makers use different scales when evaluating different subjects. Therefore, linguistic terms such as "very good" are not used universally as they rather depend on a subject of evaluation. Another important feature found by the presented study is that the 5-item Likert scale is not perceived to be equidistant by decision makers, and this conclusion is statistically highly significant. The linguistic term "very poor" was found 'wider' than other four terms.

Moreover, the study found that regional, gender, and age differences were also highly statistically significant. While respondents from Ecuador and the Czech Republic used almost identical LS-PS correspondence, respondents from France were markedly "less demanding". In addition, men provided LS-PS correspondence with higher boundary values than women in general. Similarly, older respondents provided LS-PS correspondence with higher boundary values than younger respondents.

The results of the presented study suggest that the problem of human perception of linguistic terms is much more complex than previously thought, and decision-making models with (fuzzy) linguistic variables, which employ equidistant and universal Likert scales, constitute excessive oversimplification of reality and might lead to undesired distortion of results.

We believe the findings of our study can be used as a more accurate input for models applying linguistic variables of the form of the 5-item Likert scale and more realistic linguistic evaluation of goods or services. However, more thorough research on the topic would be desirable.

References

Allen, E., & Seaman, C. (2007). Likert Scales and Data Analyses. *Quality Progress, 40*(7), 64-65.

Blaikie, N. (2003). *Analyzing Quantitative Data. From Description to Explanation*. Sage Publications, Thousand Oaks, 353. DOI: 10.4135/9781849208604.

Carrasco, R.A., Villar, P., Hornos, M.J., & Herrera-Viedma, E. (2012). A Linguistic Multi-Criteria Decision Making Model Applied to the Integration of Education Questionnaires. *International Journal of Computational Intelligence Systems, 4*(5), 946-959.

Casola, V., Preziosi, R., Rak, M., & Troiano, L.A. (2005). Reference Model for Security Level Evaluation: Policy and Fuzzy Techniques. *Journal of Universal Computer Science, 11*(1), 150-174. DOI: 10.3217/jucs-011-01-0150.

Clason, D.L., & Dormody, T.J. (1994). Analyzing data measured by individual likert-type items. *Journal of Agricultural Education, 35*(4), 31-35. DOI:10.5032/jae.1994.04031.

Creswell, J.W. (2008). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. New Jersey, United States: Pearson Education Inc.

Dawes, J. (2008). Do Data Characteristics Change According to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research, 50*(1), 61-77.

Edwards, A.L., & Kenney, K.C. (1946). A comparison of the Thurstone and Likert techniques of attitude scale construction. *Journal of Applied Psychology, 30*(1), 72-83. DOI: 10.1037/h0062418.

Hasson, D., & Arnetz, B.B. (2005). Validation and Findings Comparing VAS vs. Likert Scales for Psychosocial Measurements. *International Electronic Journal of Health Education, 8*, 178-192.

Hesketh, B., Prior, R., Gleitzman, M., & Hesketh, T. (1988). Practical Applications and Psychometric Evaluation of a Computerised Fuzzy Graphic Rating Scale. *Advances in Psychology, 56*, 425-454.

Holeček, P., & Talašová, J. (2010). FUZZME: A new software for multiple-criteria fuzzy evaluation. *Acta Universitatis Matthiae Belii ser. Mathematics, 16*, 35-51.

Joshi, A., Kale, S., Chandel, S., & Pal, D.K. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology, 7*(4), 396-403.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology, 140*, 1-55.

Lima Jr., F.R, Carvalho, G.M.R., & Carpinetti, L.C.R. (2016). A methodology based on fuzzy inference and SCOR® model for supplier performance evaluation. Gest. Prod., São Carlos, 2016 23(3), 515-534. Available from http://www.scielo.br /scielo.php?script=sci_arttext&pid=S0104-530X2016000300515&lng=en&nrm=iso. Accessed on 11 Nov. 2018. DOI: 10.1590/0104-530x2625-15.

Lin, L., & Yeh, H. (2012). Linking Consumer Perception of Store Image Using FANP. *iBusiness, 4*(1), 18-28. DOI: 10.4236/ib.2012.41003.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Science Education, 15*(5), 625-632.

Prasad, B. (2016). Using Likert Type Data in Social Science Research. Confusion, Issues and Challenges. *International Journal of Contemporary Applied Sciences, 3*(2), 36-49.

Traylor, M. (1983). Ordinal and interval scaling. *Journal of the Market Research Society, 25*(4), 297-303.

Yan, H., Zeng, X., Bruniaux, P., Chen, Y., & Zhang, X. (2017). Development of a new knowledge-based fabric recommendation system by integrating the collaborative design process and multi-criteria decision support. *Textil Research Journal, 88*(23), 2682-2698. DOI: 10.1177/0040517517729383.

Willits, F.K., Theodori, G.L. & Luloff, A.E. (2016). Another Look at Likert Scales. *Journal of Rural Social Sciences, 31*(3), 126-139.