Analyzing the influence of the rating scale for items in a questionnaire on Cronbach coefficient alpha

María Asunción Lubiano, Manuel Montenegro,
Sonia Pérez-Fernández, and María Ángeles Gil

Abstract Questionnaires are widely used in many different fields, especially in connection with human rating. Different rating scales are considered in questionnaires to base the response to their items on. The most popular scales of measurement are Likert-type ones. Other well-known rating scales to be involved in the items in a questionnaire are visual analogue, interval-valued, fuzzy linguistic and fuzzy rating scales. This paper aims to compare these five scales by means of a simulation study. The statistical tool for the comparison (actually, for the ranking) of the scales is the Cronbach index of internal consistency or reliability of a construct from a questionnaire. Percentages of advantages of the fuzzy rating scale *vs* the other ones, as well as values of the Cronbach index for some samples, are obtained and discussed.

1 Usual imprecise-valued rating scales involved in the items of a questionnaire

Questionnaires are often considered to conduct research about attitudes and human behaviour. Measurement of attitudinal and behavioural variables, especially latent variables, is facing a great challenge nowadays. Current computing developments permit advances in the measurement of the richness of human characteristics, so rating scales and tools are improving day by day bringing new and refreshing ideas to the field which attempt to improve the accuracy for making better decisions in the applied context.

Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, 33007 Oviedo, Spain
`lubiano@uniovi.es`

• This paper has been written as a tribute to our beloved and admired colleague Professor Leandro Pardo. Thank you, Leandro, for your permanent friendship and support!

In this paper, through a vast simulation study, we compare and analyze different scales of measurement by means of one of the most well-known indicators for internal consistency or reliability of the constructs: *Cronbach's alpha* [8].

In the context of survey research, a *construct* is understood to be "the abstract idea, underlying theme, or subject matter that one wishes to measure using survey questions" (see [15]). Some constructs are very simple and can be measured using only one or a few questions for which responses are numerical or dichotomous. Other constructs are more complex and may require a rather large set of questions for which responses cannot be precisely/perfectly measured/expressed.

In designing a questionnaire with such complex constructs, most of items can be formalized in terms of imprecise-valued random magnitudes and the involved rating scales for the response to such items are usually either Likert, visual analogue, interval-valued, fuzzy linguistic-valued and fuzzy-valued.

Likert Scales (LSs) [22] format consists of some scores indicating the strength of the agreement with several assertions, the Likert type items. Sometimes, these numbers are combined with or replaced by semantic expressions in terms of quantity which are, for instance, adverbs of frequency. Despite Likert Scales have been adopted for the vast majority of the social science research communities, they present some controversy and debate by social science researchers concerning several issues, among them, the nature of the response categories and the uses of the scores. Questionnaires where items comprise Likert type-items can be easily conducted. Furthermore, the options to respond to each of the questions involve some imprecision, which seems quite coherent in the context of imprecise-valued magnitudes. However, since the choice is made within a list of a few possible, anchored for the Likert options, individual differences are almost systematically overlooked. Consequently, the number of applicable techniques to statistically analyze Likert data is quite limited, and they are mostly based either on the frequencies of different 'values' or on their position in accordance with either a certain ranking or a posterior numerical encoding, so that relevant statistical information along with the inherent imprecision can be usually lost in the analysis.

On the other hand, ***Visual Analogue Scales*** (VASs) were mostly considered to overcome the limitations with ordinal discrete Likert-type scales (see [29]). VAS are not so easy-to-use, and questionnaires involving them are usually conducted by filling out either a paper-and-pencil or a computerized form, after a small training explanation showing how to proceed (since problems with subject's ability to conceptually understand the rating method itself have been reported in the literature, see [31]). VAS has a long tradition in psychological measurement. Respondents to a VAS item/scale, mark their level of agreement to a statement by indicating a position along a continuous line between two end-points, permitting an infinite number of gradations. This analogue/continuous/graphic rating aspect of the scale differentiates from others similar measures as previous mentioned Likert scale (semantic and/or numerically pointed out). VAS properly captures individual differences because the choice is made within a continuum of possible options (actually, a bounded interval). However, the choice of the single point that best represents rater's score in visual analogue scales is usually neither easy nor natural. To require a full accuracy seems rather unrealistic in connection with intrinsically imprecise variables. This statement is in line with the quote from Popper [25], in accordance with which "... Both, precision and certainty, are false ideals. They are impossible to attain, ..., it is always undesirable to make an effort to increase precision for its own sake -especially linguistic precision- since this usually leads to lack of clarity, ...: one should never try to be more precise than the problem situation demands", as it usually happens in measuring attitudes.

Single-point rating scales, like LSs and VASs, supply valuable information regarding respondents' opinion/score on a given question. However, they are limited in capturing imprecision and uncertainty of respondent answers. In case of LSs they are also limited in capturing individual differences. As it has been pointed out by Wagner *et al.* [33], "the capturing of respondents' uncertainty requires the development of more suitable scales...".

Aiming to capture imprecision/uncertainty in responding to questions related to intrinsically imprecise magnitudes, Themistocleous *et al.* [30] highlighted that ***Interval-Valued Scales*** (IVSs) allow the respondent the "choice of an interval when providing a response by positioning an ellipse on a straight line with polar adjectives on its two ends" (see also Wagner *et al.* [33]). Consequently, items with interval-valued responses offer richer and more complex information compared with single-point rating scales and provide researchers with more insights regarding respondent perceptions as well as the imprecision/uncertainty of their responses which is expected to increase the reliability of constructs in the questionnaire. Furthermore, as IVSs do not prefix the intervals to be chosen, respondents are completely free in expressing their answers, so they can capture individual differences. In addition to reflect inherent imprecision associated with most of latent variables and constructs in questionnaires, IVSs are fully suitable either to model magnitudes related to ranges or, more generally, interval-valued symbolic data.

On the other hand, *fuzzy scales* were introduced to establishing a bridge between strongly defined measurements, as the VASs or the numerically encoded LSs, and weakly defined measurements used in behavioral sciences as the Likert or the semantic differential.

Fuzzy Linguistic Scales (FLSs), associated with the so-called fuzzy linguistic variables, were stated by Zadeh [35] as a flexible alternative to the numerical encoding of Likert and semantic differential scales. In fact, the numerical encoding does not take into account the essential imprecision accompanying 'values' of most of the variables in attitudinal studies, whereas the fuzzy linguistic encoding (partially) does, although individual differences cannot be grasped either by LSs or by FLSs.

Aiming to overcome the last drawback, Hesketh *et al.* [20] and [21], introduced ***Fuzzy Rating Scales*** (FRSs) as an extension of the other mentioned scales. To motivate, justify and support introducing FRS's, they argued that "A perennial issue in psychological assessment has been the extent to which differences in psychological test scores are a function of genuine individual differences rather than differences imposed as it happens with visual analogue scales, or obscured as it happens with Likert or semantic differential scales by the constraints of the measurement procedures." FRS's adapt the semantic differential so that a preferred point on a given interval (with anchored end points), along with latitudes of acceptance on either side should be indicated by the respondent. [19] allowed the preferred 'point' to be a subinterval within the original given interval. The preferred point/subinterval determines the 'core' (1-level) of the fuzzy assessment (i.e., the value or interval of values that are fully compatible with respondent's rating). When the core is enlarged with the latitudes of acceptance, one gets the 'support' (topological closure of the 0-level) of the fuzzy assessment (i.e., the interval of values that are compatible to some extent with respondent's rating). And the choice of core and support is completely free, so no list of possible responses is prefixed.

2 Comparing rating scales through Cronbach alpha

The use of scales like IVSs, FLSs and FRSs in connection with questionnaires is relatively new in contrast to that of either discrete or continuous single-point scales like LSs and VASs, respectively. As noticed by Ellerby *et al.* [14] and Lubiano *et al.* [23], the incorporation of these scales suggests to apply and mainly to develop methodologies for the statistical analysis of interval- and fuzzy-valued responses, as well as to compare the new scales with the single-point ones.

Regarding the approaches to the statistical analysis of interval- and fuzzy-valued data, several studies about can be found in the literature of the last two decades. Although, at present not all the problems and methods to statistically analyze real-valued data have been extended to deal either with interval-valued or with fuzzy-valued data, some interesting ones do (see, for instance, [2, 3, 4, 5, 6, 11, 12, 13, 16, 17, 18, 23, 24, 26, 28]).

This paper focuses on the comparison between the above mentioned rating scales. The comparison is to be based on examining the behavior of the well-known Cronbach coefficient of internal consistency/reliability of constructs in a questionnaire by means of simulation developments. Along this simulation study, LS will be identified with the numerically encoding of the Likert type scale and the Cronbach α will be given as follows:

Definition 1. Given a construct involving K items and the response of the ith respondent ($i = 1, \ldots, n$) to the jth item ($j = 1, \ldots, K$) being denoted by

– x_{ij} if a single-point-valued scale is considered,
– x_{ij} if an interval-valued scale is considered
– $\widetilde{x}_{ij}$ if a fuzzy-valued scale is considered,

the ***Cronbach*** $\boldsymbol{\alpha}$ is given by

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{j=1}^{K} s_j^2}{s_{\text{total}}^2}\right),$$

where s_j^2 is the sample variance of the responses to item j, and s_{total}^2 is the variance of all the responses to the items involved in the construct, the variances being defined as the Fréchet ones w.r.t. the Euclidean distance in $\mathbb{R}$ for the single-point scales, the Vitale δ_2-metric [32] for the interval-valued scale and the Diamond and Kloeden ρ_2-metric [10] for the fuzzy-valued scales, that is,

$$s_j^2 = \begin{cases} \displaystyle\sum_{i=1}^{n} \frac{\left[x_{ij} - \overline{x}_j\right]^2}{n} & \text{for LS/VAS} \\ \displaystyle\sum_{i=1}^{n} \frac{\left[\inf \mathsf{x}_{ij} - \overline{\inf \mathsf{x}}_j\right]^2 + \left[\sup \mathsf{x}_{ij} - \overline{\sup \mathsf{x}}_j\right]^2}{2n} & \text{for IVS} \\ \displaystyle\sum_{i=1}^{n} \int_{[0,1]} \frac{\left[\inf(\widetilde{x}_{ij})_v - \overline{\inf(\widetilde{x}_v)}_j\right]^2 + \left[\sup(\widetilde{x}_{ij})_v - \overline{\sup(\widetilde{x}_v)}_j\right]^2}{2n}\, dv & \text{for FLS/FRS} \end{cases}$$

$$s_{\text{total}}^2 = \begin{cases} \displaystyle\sum_{j=1}^{K}\sum_{i=1}^{n} \frac{\left[x_{ij} - \overline{x}\right]^2}{nK} & \text{for LS/VAS} \\ \displaystyle\sum_{j=1}^{K}\sum_{i=1}^{n} \frac{\left[\inf \mathsf{x}_{ij} - \overline{\inf \mathsf{x}}\right]^2 + \left[\sup \mathsf{x}_{ij} - \overline{\sup \mathsf{x}}\right]^2}{2nK} & \text{for IVS} \\ \displaystyle\sum_{j=1}^{K}\sum_{i=1}^{n} \int_{[0,1]} \frac{\left[\inf(\widetilde{x}_{ij})_v - \overline{\inf(\widetilde{x}_v)}\right]^2 + \left[\sup(\widetilde{x}_{ij})_v - \overline{\sup(\widetilde{x}_v)}\right]^2}{2nK}\, dv & \text{for FLS/FRS} \end{cases}$$

where $\widetilde{x}_v = \{t \in \mathbb{R} : \widetilde{x}(t) \geq v\}$.

It should be remarked in connection with the value of α for fuzzy-valued data, that they are scarcely influenced by the shape chosen for such data (see Lubiano *et al.* [23]), so to assume this shape is trapezoidal does not mean a real constraint.

In comparing different rating scales through α general conclusions cannot be drawn, but one can get majority trends by means of simulations.

2.1 Simulation of FRS-based responses and suggested links with responses to other rating scales

Since there are not yet suitable realistic models for the distribution of the random mechanisms generating fuzzy responses/data, the simulation process is not an immediate and standard one. However, in previous papers dealing with the statistical analysis of fuzzy data simulation procedures have been introduced (see, for instance, [9, 27]). For purposes of analyzing reliability of constructs some relationships between responses to items should be added (see Lubiano *et al.* [23]).

By combining the previous procedures we will alternatively denote

$$\mathrm{Tra}(a,b,c,d) = \mathrm{Tra}\langle x_1, x_2, x_3, x_4\rangle,$$

where

$$x_1 = (b+c)/2,\ x_2 = (c-b)/2,\ x_3 = b-a,\ x_4 = d-c,$$

(see Figure 1 to illustrate the double notation). The simulation process will generate the 4-tuple (x_1, x_2, x_3, x_4) in accordance with some guideliness to be now explained.

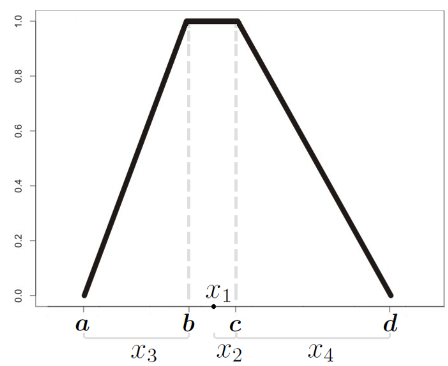

Fig. 1 A 4-tuple (x_1, x_2, x_3, x_4) generated from the simulation process, and the associated trapezoidal fuzzy datum

To each generated 4-tuple $(x_1, x_2,\ x_3, x_4)$ we associate the trapezoidal fuzzy datum $\mathrm{Tra}\langle x_1, x_2, x_3, x_4\rangle = \mathrm{Tra}(x_1 - x_2 - x_3, x_1 - x_2, x_1 + x_2, x_1 + x_2 + x_4)$.

By inspiring the simulation process in most of the already known real-life examples, fuzzy data will be generated as follows:

– 5% (or, more generally, $100 \cdot \omega_1\%$) of the data have been obtained by first considering a simulation from a simple random sample of size 4 from a beta $\beta(p, q)$ distribution, the ordered 4-tuple, and finally computing the values of the x_i. The values of p and q have been assumed to be $p = q = 1$. The values from the beta distribution should be re-scaled and translated to the reference interval $[l_0, u_0]$ for the considered FRS.
– 35% (or, more generally, $100 \cdot \omega_2\%$) of the data have been obtained considering a simulation of four random variables $X_i = (u_0 - l_0) \cdot Y_i + l_0$ as follows:
  $Y_1 \sim \beta(p, q)$,
  $Y_2 \sim \mathrm{Uniform}\big[0, \min\{1/10, Y_1, 1 - Y_1\}\big]$,
  $Y_3 \sim \mathrm{Uniform}\big[0, \min\{1/5, Y_1 - Y_2\}\big]$,
  $Y_4 \sim \mathrm{Uniform}\big[0, \min\{1/5, 1 - Y_1 - Y_2\}\big]$.

– 60% (or, more generally, $100 \cdot \omega_3\%$) of the data have been obtained considering a simulation of four random variables $X_i = (u_0 - l_0) \cdot Y_i + l_0$ as follows:

$$Y_1 \sim \beta(p, q),$$
$$Y_2 \sim \begin{cases} \mathrm{Exp}(200) & \text{if } Y_1 \in [0.25, 0.75] \\ \mathrm{Exp}(100 + 4\,Y_1) & \text{if } Y_1 < 0.25 \\ \mathrm{Exp}(500 - 4\,Y_1) & \text{otherwise} \end{cases}$$
$$Y_3 \sim \begin{cases} \gamma(4, 100) & \text{if } Y_1 - Y_2 \geq 0.25 \\ \gamma(4, 100 + 4\,Y_1) & \text{otherwise} \end{cases}$$
$$Y_4 \sim \begin{cases} \gamma(4, 100) & \text{if } Y_1 + Y_2 \geq 0.25 \\ \gamma(4, 500 - 4\,Y_1) & \text{otherwise.} \end{cases}$$

To add the possible relationship between responses to items in analyzing reliability, a large sample of $n = 500$ FRS-type data for each of a large number of items, $K = 100$, is to be simulated in accordance with the above described generation procedure. This process will provide with an 'auxiliary sample' from which we will later select data for other choices of n and K and transform them to mimic a certain (linear) dependence. To generate the 500×100 data we proceed as follows:

S1. A sample of 500 FRS-type data $(\widetilde{x}_1^*, \ldots, \widetilde{x}_{500}^*)$, the reference interval of the FRS being $[l_0, u_0]$, are first simulated as the 'auxiliary sample'.

S2. To mimic the desirable high correlation between the responses from a respondent to different 100 items, for any item j $(j = 1, \ldots, 100)$

- a pair (γ_j, δ_j) is considered so that γ_j is generated at random from a uniform distribution in $[0, 1]$ and δ_j is generated from a standard normal distribution;
- the response of the i-th respondent $(i = 1, \ldots, 500)$ to the j-th item is assumed to be given by $\widetilde{x}_{ij} = \gamma_j \cdot \widetilde{x}_i^* + \delta_j + \varepsilon_{ij}$, with ε_{ij} being generated at random from a standard normal distribution;
- in case any $\widetilde{x}_{ij}$ is not fully included within interval $[l_0, u_0]$, the response is appropriately truncated.

Once we get the simulated largest data set including 500×100 fuzzy data, we choose at random and stepwise $n = 450$ from the former 500 respondents, $n = 400$ from the preceding selected 450 respondents, and so on. Analogously, we choose at random and stepwise $K = 50$ from the former 100 items, $K = 40$ from the preceding selected 50 items, and so on. To be realistic, in the studies in the paper we will constrain K to take on values up to 50.

In what concerns the links between the FRS-based responses and the ones based on the other rating scales, we will consider some reasonable and realistic ones as follows:

- The numerically encoded r-point Likert scale usually considered is 1, 2, …, r, but to compare it with the FRS the values 1 to r should be rescaled in accordance with the reference interval $[l_0, u_0]$, so that $L_i = l_0 + (u_0 - l_0) \cdot (i-1)/(r-1)$, for $i \in \{1, \ldots, r\}$. The link between FRS and the numerically encoded LS will be the one associated with the minimum ρ_2-distance criterion, i.e., if $\widetilde{x} = \mathrm{Tra}(a, b, c, d)$ is the available FRS-valued response
$$\widetilde{x} \leftrightarrow \mathrm{L}(\widetilde{x}) = \arg \min_{L_i,\, i \in \{1,\ldots,r\}} \left[2 \cdot L_i^2 - (a+b+c+d) \cdot L_i\right],$$
corresponding to the L_i being closer to the 'central point' $(a+b+c+d)/4$.
- In mimicking the connection between FRS and VAS responses for the same respondent, a reasonable link is the one associated with a suitable 'defuzzification' process like the one introduced in [34], which for a trapezoidal response $\widetilde{x} = \mathrm{Tra}(a, b, c, d)$ is such that
$$\widetilde{x} \leftrightarrow \mathrm{VA}(\widetilde{x}) = \frac{a+b+c+d}{4}.$$
- In mimicking the connection between FRS and IVS responses for the same respondent, a possible link is the one associated with the .5-level of the fuzzy response, which for a trapezoidal response $\widetilde{x} = \mathrm{Tra}(a, b, c, d)$ is such that
$$\widetilde{x} \leftrightarrow \mathrm{IV}(\widetilde{x}) = (\widetilde{x})_{.5} = \left[\frac{a+b}{2}, \frac{c+d}{2}\right].$$
- Finally, the connection between FRS and FLS responses for the same respondent, a possible link is to consider the numerical 'Likertization' in the first stated connection and to consider later, for instance, the linguistic hierarchy of r labels (see, for instance, Cordón *et al.* [7]). Figure 2 graphically displays the one associated with a 5-point Likert scale when $[l_0, u_0] = [0, 100]$. At this point, it should be emphasized that the choice of the linguistic fuzzification scarcely affects the value of the Cronbach α.

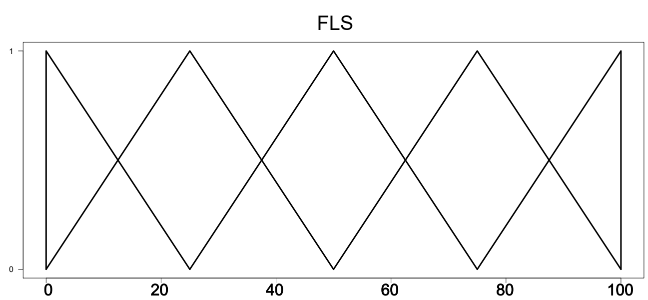

Fig. 2 Frequently used fuzzy linguistic encoding of 5-point Likert scales

2.2 Comparison of rating scales through percentages of greater values of α

The comparative studies in this paper consider the reference interval to be $[l_0, u_0] = [0, 1]$ and the involved Likert scale to be the 5-point one in Figure 2.

For different choices of n (number of respondents) and K (number of items), 1000 samples of $n \times K$ FRS-based data have been generated, and later linked, by means of the process described in Section 2.1. Later, percentages of samples for which Cronbach's α of the FRS-based data is greater than that of the other rating scales are computed and collected. Table 1 shows a few choices of K up to 50, albeit outputs for larger values are rather similar.

Table 1 Percentages of simulated samples for which Cronbach's α of the FRS are greater than that of the other rating scales for different choices of n (number of respondents) and K (number of items)

n & K choices		$\alpha_{\text{FRS}} > \alpha_{\text{IVS}}$	$\alpha_{\text{FRS}} > \alpha_{\text{VAS}}$	$\alpha_{\text{FRS}} > \alpha_{\text{LS}}$	$\alpha_{\text{FRS}} > \alpha_{\text{FLS}}$
n=300	K=50	100	100	100	100
	K=30	100	100	100	100
	K=20	100	100	100	100
	K=10	100	100	99.9	100
	K=5	99.4	99.4	95.7	97.9
n=100	K=50	98.2	100	100	100
	K=30	99.6	100	100	100
	K=20	100	100	100	100
	K=10	99.8	99.8	97.8	99.6
	K=5	98	97.9	87.3	93.4
n=50	K=50	94.6	100	100	100
	K=30	98.2	100	100	100
	K=20	98.9	99.8	99.5	100
	K=10	98.8	99	92.5	97.6
	K=5	95.2	95.26	80.6	87.8
n=30	K=50	90.1	99.9	99.7	100
	K=30	94.9	99.7	99.6	100
	K=20	96.5	99.1	97.8	99.8
	K=10	96.5	98.2	89.5	95.4
	K=5	91.2	91.9	75.4	83.1

Consequently, in getting a larger internal consistency/reliability of a construct, majority trends support the almost general superiority of the FRS with respect to the other rating scales.

2.3 Comparison of rating scales through values of α

In addition to the advantage of the FRS w.r.t. the other ones in terms of the percentages of greater reliability, it would be interesting to examine whether such advantage is also clear in terms of the values of Cronbach coefficient. Table 2 gathers values of Cronbach's α of the five rating scales for a sample of size of $n \times K$ FRS-based data generated, and later linked, by means of the process described in Section 2.1.

Table 2 Values of Cronbach's α for the rating scales and a sample of size $n \times K$ for different choices of n and K

n & K choices		α_{FRS}	α_{IVS}	α_{VAS}	α_{LS}	α_{FLS}
n=300	K=50	.9187	.9185	.9161	.9083	.9047
	K=30	.8624	.8620	.8584	.8447	.8387
	K=20	.8405	.8401	.8361	.8215	.8159
	K=10	.7533	.7527	.7476	.7305	.7228
	K=5	.6779	.6773	.6714	.6511	.6385
n=100	K=50	.9134	.9129	.9093	.9010	.8975
	K=30	.8489	.8482	.8424	.8269	.8196
	K=20	.8204	.8196	.8129	.7951	.7878
	K=10	.7252	.7239	.7154	.7048	.6962
	K=5	.6774	.6765	.6697	.6565	.642
n=50	K=50	.9142	.9135	.9101	.9026	.8999
	K=30	.8303	.8289	.8214	.8082	.8012
	K=20	.8037	.8022	.7941	.7805	.7742
	K=10	.6919	.6895	.6788	.6834	.6753
	K=5	.674	.6725	.6673	.6772	.6596
n=30	K=50	.9207	.9205	.9188	.9107	.9080
	K=30	.8493	.8489	.8455	.8321	.8235
	K=20	.8131	.8127	.8085	.7959	.7861
	K=10	.7103	.7096	.7039	.7103	.6969
	K=5	.7329	.7325	.7301	.7332	.7149

In a similar way, a graphical comparison is displayed in Figure 3 for a sample in which $n = 100$, corroborating that the ranking with respect to the reliability of constructs is FRS–IVS–VAS–LS–FLS, the difference between FRS and IVS being really a minor one.

The research in this paper can be complemented with the comparisons based on alternative tools or indicators, like those for the validation of questionnaires that are closely connected with divergences, one of the highest research interest of our tributed colleague (see, for instance, [1]), as well other possible links between scales. This complementary analysis will help also in making decisions on the number of items for achieving a given reliability/indicator value, on the convenient scale to choose, and so on.

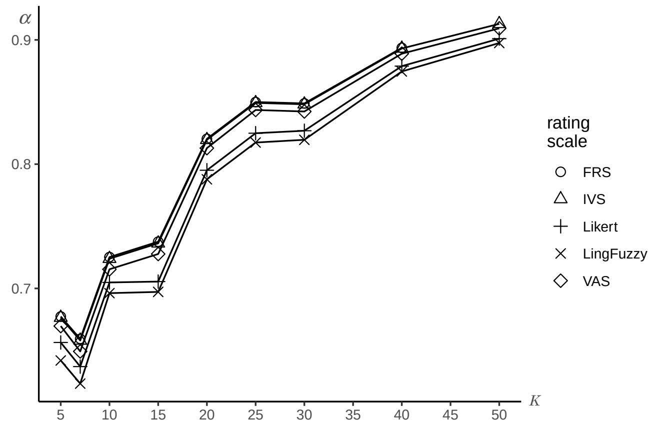

Fig. 3 Evolution of values of Cronbach's α

Acknowledgements This research has been partially supported by the Spanish Ministry of Science and Innovation PID2019-104486GB-I00, which is gratefully acknowledged.

References

1. Balakrishnan N (2009) Methods and Applications of Statistics in the Life and Health Sciences. J Wiley, New York
2. Billard L, Diday E (2006) Descriptive statistics for interval-valued observations in the presence of rules. Comp Stat 21(2):187–210
3. Blanco-Fernández Á, Casals MR, Colubi A, Corral N, García-Bárzana M, Gil MÁ, González-Rodríguez G, López MT, Lubiano MA, Montenegro M, Ramos-Guajardo AB, De la Rosa de Sáa S, Sinova B (2014) A distance-based statistical analysis of fuzzy number-valued data. Int J Approx Reas 55(7):1487–1501; Rejoinder. Int J Approx Reas 55(7):1601–1605
4. Carvalho FDAT, Lima Neto EDA, da Silva KCF (2021) A clusterwise nonlinear regression algorithm for interval-valued data. Inform Sci 555:357–385
5. Colubi A, González-Rodríguez G, Gil MÁ, Trutschnig W (2011) Nonparametric criteria for supervised classification of fuzzy data. Int J Approx Reas 52:1272–1282
6. Coppi R, D'Urso P, Giordani P (2012) Fuzzy and possibilistic clustering for fuzzy data. Comp Stat Data Anal 56(4):915–927
7. Cordón O, Herrera F, Zwir I (2002) Linguistic modeling by hierarchical systems of linguistic rules. IEEE Trans Fuzzy Syst 10(1):2–20
8. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16(3):297–334
9. De la Rosa de Sáa S, Gil MÁ, González-Rodríguez G, López MT, Lubiano MA (2015) Fuzzy rating scale-based questionnaires and their statistical analysis. IEEE Trans Fuzzy Syst 23(1):111–126
10. Diamond P, Kloeden P (1990) Metric spaces of fuzzy sets. Fuzzy Sets Syst 35:241–249
11. D'Urso P, De Giovanni L, Massari R (2015) Trimmed fuzzy clustering for interval-valued data. Adv Data Anal Class 8(1):21–40
12. D'Urso P, Gil MÁ (2017) Fuzzy data analysis and classification. Special issue in memoriam of Professor Lotfi A. Zadeh, father of fuzzy logic. Adv Data Anal Class 11(4):645–657

13. D'Urso P, Giordani P (2004) A least squares approach to principal component analysis for interval valued data. Chem Intel Lab Syst 70(2):179–192
14. Ellerby Z, Wagner C, Broomell S (2021) Capturing richer information–On establishing the validity of an interval-valued survey response mode. Behav Res Meth, in press
15. Encyclopedia of Survey Research Methods (2008) Lavrakas PJ (ed). SAGE Pub, Inc, Thousand Oaks, CA
16. García-Bárzana M, Ramos-Guajardo AB, Colubi A, Kontoghiorghes EJ (2020) Multiple linear regression models for random intervals: a set arithmetic approach. Comp Stat 35(2):755–773
17. Gil MÁ, González-Rodrguez G, Colubi A, Montenegro, M (2007) Testing linear independence in linear models with interval-valued data. Comp Stat Data Anal 51(6):3002–3015
18. González-Rodríguez G, Colubi A, Gil, MÁ (2012) Fuzzy data treated as functional data. A one-way ANOVA test approach. Comp Stat Data Anal 56(4):943–955
19. Hesketh T, Hesketh B (1994) Computerized fuzzy ratings: The concept of a fuzzy class. Behav Res Meth Instr Comp 26(3):272–277
20. Hesketh B, Pryor R, Gleitzman M, Hesketh T (1988a) Practical applications and psychometric evaluation of a computerised fuzzy graphic rating scale. In: Zétényi T (ed) Fuzzy Sets in Psychology. Advances in Psychology Series, 56(C):425–454. North-Holland/Elsevier, Amsterdam
21. Hesketh T, Pryor R, Hesketh B (1988b) An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences. Int J Man-Mach Stud 29(1):21–35
22. Likert R (1932) A technique for the measurement of attitudes. Arch Psychol 22:140–155
23. Lubiano MA, García-Izquierdo AL, Gil MÁ (2021) Fuzzy rating scales: Does internal consistency of ameasurement scale benefit from coping with imprecision and individual differences in psychological rating? Inform Sci 550:91–108
24. Lubiano MA, Montenegro M, Sinova B, De la Rosa de Sáa S, Gil MÁ (2016) Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications. Eur J Oper Res 251(3):918–929
25. Popper K (1974) Autobiography by Karl Popper - The Philosophy of Karl Popper. The Lib Living Phil, Inc., La Salle, IL
26. Sinova B (2016) M-estimators of location for interval-valued random elements. Chem Intel Lab Syst 156: 115—127
27. Sinova B, Gil MÁ, Colubi A, Van Aelst S (2012) The median of a random fuzzy number. The 1-norm distance approach. Fuzzy Sets Syst 200:99–115
28. Sinova B, Van Aelst S, Terán P (2021) M-estimators and trimmed means: from Hilbert-valued to fuzzy set-valued data. Adv Data Anal Class 15(2):267–288
29. Sung YT, Wu JS (2018) The Visual Analogue Scale for rating, ranking and paired-comparison (VAS-RRP): A new technique for psychological measurement. Beh Res Meth 50:1694–1715
30. Themistocleous C, Pagiaslis A, Smith A, Wagner C (2019) A comparison of scale attributes between interval-valued and semantic differential scales. Int J Market Res 61(1):394–407
31. Visual analog scales (2018) In: Frey BB (ed) The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation, Vol 4. SAGE Pub, Inc, Thousand Oaks, CA
32. Vitale RA (1985) Metrics for compact, convex sets. J Approx Theor 45:280–287
33. Wagner C, Miller S, Garibaldi JM, Anderson DT, Havens TC (2015) From interval-valued data to general type-2 fuzzy sets. IEEE Trans Fuzzy Syst 23:248–269
34. Yager RR (1981) A procedure for ordering fuzzy subsets of the unit interval. Inform Sci:143–161
35. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. Part 1. Inform Sci 8:199–249; Part 2. Inform Sci 8:301–353; Part 3. Inform Sci 9:43–80