Original scientific paper

UDC: 159.9.072.59:3

Received: April, 21, 2026.
Revised: June, 30, 2026.
Accepted: July, 07, 2026.

10.23947/2334-8496-2026-14-2-357-372

Check for updates

A New Scale in Education (Social Science)

Sinisa Opic[1*]

[1] University of Zagreb, Faculty of Teacher Education, Croatia, e-mail: sinisa.opic@ufzg.hr

Abstract: This paper introduces a new scale in the social sciences that combines two types of measurements: a binary choice and a relative interval measurement consisting of 10+10 intervals. The scale's relative interval nature allows for differentiation between intervals of 10±10, expressed based on the principle of deciles. This approach provides a more nuanced differentiation between degrees compared to traditional Likert scales. The paper also discusses a coding method and includes graphical representations of the data. An important aspect of the scale is its capacity to categorize responses into positive and negative attitudes (NEGATT and POSATT), enabling separate descriptive and inferential statistical analyses.

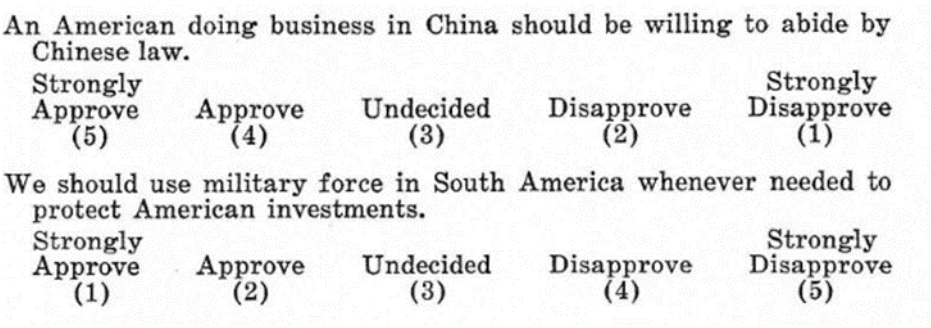
Keywords: *binary choice, intervals, measurements, new scale, social science.*

Introduction - In memory of Rensis Likert, back in 1932

Social sciences, with their habitus and the complexity of the phenomena being investigated, are extremely complex in terms of the generalizations and assumptions that follow from research results. A series of challenges arise, such as the influence of several factors on the object of measurement (dependent/dependent variables) and the weak power of their control, the complexity/entanglement of the bio-psycho-social structure of a person and its manifest features in social phenomena and relationships, the problem of objectivity, bias, incoherence, implausibility, non-nomothetic, reliability, etc. From a methodological perspective, these challenges can make research in the social sciences uniquely complex. The lack of fundamental research is a key reason why social sciences lag behind medical, technical, and natural sciences, frequently struggling to keep pace with modern development trends. One could argue that social sciences are inherently challenging to explain objectively and reliably. Are they scientifically trustworthy? A significant advancement in establishing the scientific credibility of social sciences has been the development of a dominant scale for measuring attitudes: the Likert scale.

American social psychologist Rensis Likert, who was a faculty member at the University of Michigan from 1946 to 1970, developed a rating scale for measuring public attitudes toward international relations as part of his doctoral dissertation in 1932. The primary goal of this scale was to create a scientific method for assessing attitudes. Since its creation, the scale has garnered significant interest and has been applied in various scientific disciplines.

*Corresponding author: sinisa.opic@ufzg.hr


An American doing business in China should be willing to abide by Chinese law.

Strongly Approve (5)	Approve (4)	Undecided (3)	Disapprove (2)	Strongly Disapprove (1)

We should use military force in South America whenever needed to protect American investments.

Strongly Approve (1)	Approve (2)	Undecided (3)	Disapprove (4)	Strongly Disapprove (5)

Figure 1. The original form of the Likert scale looks like this (Likert, 1932, p. 20)

Scholarly debate about the scale has continued for over half a century, with one of the key disputes being:

Is the Likert scale exclusively ordinal?

If it is exclusively ordinal, then we cannot use parametric, but exclusively non-parametric tests (Bishop and Herron, 2015). However, in the social sciences, we often use parametric tests on an ordinal type scale where they cannot be used. Accordingly, it is crucial to understand the implications of this misuse, as it can lead to misleading results and interpretations. Researchers should be cautious and ensure they select the appropriate statistical methods that align with the scale of measurement to maintain the integrity of their findings. Accordingly, Jakobsson (2004) analyzed 166 full-length articles from the 2003 editions of Cancer Nursing, Scandinavian Journal of Caring Sciences and Nursing. Studies were reviewed for their use of ordinal data. The author concluded that ordinal scales were utilized in approximately one-third of the articles. However, only about half of the analyses of the ordinal data were conducted using the most appropriate statistical approaches (table 1).

Table 1. *Ordinal data in the article (Ibidem, 438)*

Journal	n	Appropriate presentation (%)	Appropriate analysis (%)
Cancer Nurs	17	53	38
Scand J Caring Sci	16	69	93
Nurs Res	18	28	38
Total	51	49	57

Although Jakobsson's (2004) study focused on medicine and nursing, a broader investigation in the social sciences could reveal interesting findings about authors' use of parametric tests on data collected using ordinal scales. This is because analyzing data collected on an ordinal scale using parametric tests is inappropriate (Pett, 1997; Lewis and Traill, 1999; Selimović, Opić and Selimović, 2023). On the other hand, Carifio and Perla (2007) discussed urban legends and ten myths about the Likert scale, providing logical counterarguments as antidotes. According to the authors, Myth 6 states that because the Likert scale is ordinal, only non-parametric tests can be used with it. They argue that this belief is incorrect, citing a Monte Carlo study that indicates that the F-test/ANOVA is highly robust even when the assumptions of using an interval scale are not met (Glass, Peckham, and Sanders, 1972). Supporting this interpretation is the viewpoint expressed by Harris (1975):

> "A number of authors… have pointed out that statistical conclusions are valid whenever the distributions of numbers from which the data are sampled meet the assumptions (typically, normality and homogeneity of variance) used to derive the particular techniques being applied, irrespective of the measurement process which generated those numbers. Moreover, the validity of parametric statistics… is often affected very little by even relatively gross departures from these assumptions"

Allen and Seaman (2007) indicate that combining Likert scales into indices contributes to the value and variability of the data. In this case, if the assumption of normality of the distribution is met, parametric

statistics can be used. Tanujaya, Prahmana, and Mumu (2022) believe that the Likert scale becomes an interval scale when data generated are composite scores derived from multiple items. They also state that the Likert scale can also be used for parametric tests such as correlation, regression, z-test, t-test, ANOVA, Structural equation modeling, and Factor analysis (also sum, mean, and standard deviation) (p. 97). Guerra, Gidel, and Vezzetti (2016) avoid giving a definitive answer to whether parametric tests can be applied to Likert scale data, referencing Adams, Fagot, and Robinson (1965):

> "Nothing is wrong per se in applying any statistical operation to measurements of given scale, but what may be wrong, depending on what is said about the results of these applications, is that the statement about them will not be empirically meaningful or else that it is not scientifically significant" (p. 100).

What is the optimal number of points on a Likert scale?

The five-point Likert scale is the most commonly used, but there are varying opinions on the advantages and disadvantages of the number of categories or levels in the scale. According to Yaska and Nuhu (2024), who analyzed measures of central tendency and variability, the 5-point Likert scale is the optimal choice when compared to a 4-point Likert scale. Chang (1994) compared the validity and reliability of 4-point and 6-point Likert scales. In an empirical approach, different measurement models were fit by confirmatory factor analyzes of a multitrait-multimethod covariance matrix. The results showed lower reliability (heterotrait-monomethod coefficients) with the 6-point scale than with the 4-point scale, but when it comes to validity, there was no difference between the 4-point and 6-point scales.

Conjunction results of research on the influence of the number of scale points on measurement capabilities (reliability) are contradictory. For example, Brown, Widing and Coulter (1991) analyze the SOCO scale and conclude that "(…) the reduction in scale points and accompanying anchors did not lead to any reduction in the scale's measurement capabilities" (p. 350).

Research shows similar results about how the number of scale levels does not substantially affect psychometric properties (reliability, validity): Finstad (2010), Nunnally and Bernstein, (1994). However, McCallum, Keith and Wiebe (1988), Boote (1981), Hernández, Muñiz and García Cueto (2000), Moreno, Martínez and Muñiz (2004), Muñiz, García-Cueto and Lozano (2005) show the opposite, that is, how the number of levels (points) on a scale affects psychometric properties. Lozano, García-Cueto, and Muñiz (2008) found that increasing the number of points on a scale enhances both validity and reliability. They observed that when a scale has fewer than four points, both reliability and validity tend to decrease. The authors concluded that while increasing the scale beyond seven points does not significantly improve validity and reliability, the optimal number of response options lies between four and seven. Leung (2011) investigated some psychometric properties of 4,5,6, and 11-point Likert scales on a sample of 1217 students in Macau. There were no significant differences found in the internal structure regarding means, standard deviations, item-item correlations, item-total correlations, Cronbach's alpha, or factor loadings. The results suggest that increasing the number of scale points helps reduce skewness, with the 11-point scale (ranging from 0 to 10) showing the lowest kurtosis and being closest to a normal distribution. Only the 6- and 11-point scales align with normal distributions based on Kolmogorov-Smirnov and Shapiro-Wilk tests. The findings on predictive validity are inconclusive. This article recommends adopting an 11-point scale, as it enhances sensitivity and approximates interval-level measurement and normality.

The optimal number of categories on the Likert scale in the literature varies. It ranges from the dichotomous feature (2) to 3 categories, then 4 with the argument of avoiding anchoring neutral values (or without it), 5 (the most common category in social sciences), 6, 7, 9, 11 categories, and more. Moreover, Garner (1960) concludes that " information transmission measures, after correction for sampling bias, show a small but definite increase in discrimination up to the 20 categories used" (p 352).

An intriguing research study by Miller (1994) demonstrates:

"What about the magical number seven? What about the seven wonders of the world, the seven seas, the seven deadly sins, the seven daughters of Atlas in the Pleiades, the seven ages of man, the seven levels of hell, the seven primary colors, the seven notes of the musical scale, and the seven days of the week? What about the seven-point rating scale, the seven categories for absolute judgment, the seven objects in the span of attention, and the seven digits in the span of immediate memory? For the present, I propose to withhold judgment. Perhaps there is something deep and profound behind all these

sevens, something just calling out for us to discover it. But I suspect that it is only a pernicious, Pythagorean coincidence" (p. 351).

The author suggests, albeit in a poetic manner, that a scale of 7 points, with a margin of plus or minus two, is optimal. Additionally, the findings of Green and Rao (1970) indicate that expanding the scale beyond 7 points (response categories) is unnecessary.

Preston and Colman (2000) examine the reliability, validity, and discriminative power of various point scales. The results showed that scales with two, three, and four response categories performed poorly across various indices of reliability, validity, and discriminating power. In contrast, scales with more categories, up to seven, showed significantly better results. While internal consistency remained similar across all scales, test-retest reliability tended to decrease for scales with more than 10 response categories. Respondents expressed the highest preference for the 10-point scale, followed closely by the 7-point and 9-point scales.

Schrum, Johnson, Ghuy, Gombolay (2002) suggest a scale of 5-9 points as a rule of thumb; less than 5 or more than 9 points cause a lack of precision. Aybek and Toraman (2022) compare 3-point, 5-point, and 7-point Likert scales. They conclude that the 5-point scale provides greater reliability than the 3-point scale, while there is no significant difference in reliability between the 5-point and 7-point scales. Therefore, they recommend using the 5-point scale as the best option. De Vellis (2003) states that a scale with a neutral value can lead to ambiguity in the interpretation of data (strongly disagree, disagree, *somewhat disagree, somewhat agree,* agree, and strongly agree). The authors Pearson, Lacombe and Khatun (2024) conclude the opposite, outlining that overlooking/ignoring a "neutral" category can result in flawed inferences, which can impede the advancement of personal finance research (can lead to biased estimates).

Increasing the number of points on a scale can lead to greater variability, particularly in small sample sizes. This may violate the assumptions of a normal distribution (such as the coefficient of variability) and can also impact the asymmetry of sampling distributions. While a higher number of scale points does not always lead to increased variability, especially when respondents' attitudes are highly polarized, it can result in a U-shaped distribution. Another significant issue with expanding the Likert scale is that it becomes difficult for individuals to clearly distinguish between certain points or categories in a semantic context. As the number of scale points increases, cognitive differentiation becomes problematic.

The limitations of the Likert scale can be summarized as follows (Jamieson, 2000):

1. The response categories in Likert scales are ordered, but the intervals between the values should not be assumed to be equal.
2. A Likert scale is inappropriate for parametric tests (based on the mean and variance/standard deviation).
3. Treating ordinal scales as interval scales can be very controversial.
4. It has become a common practice to treat Likert-type categories as if they represent interval-level measurement.

We can also expand this to:

5. Giving socially desirable answers (especially when it comes to semantic categories).
6. Tending to give middle/neutral values (if the number of degrees on the scale is odd, id est. if it has a neutral value - the midpoint of "neither agree nor disagree"(Chyung et al., 2017).
7. Confusing semantic differences of opposing categories (e.g., category 2 – I partially disagree vs category 4 – I partially agree). Statistically, it is a huge difference, but in a normative linguistic statement, this is almost the same, because if I partially agree then I also partially disagree. How much is partially positive greater than partially negative?
8. Unnecessary avoidance of non-parametric statistics - avoidance of categorical dichotomous/ trichotomous/ polytomous variables (not rank). In medicine, for example using procedures like Odds ratio, probability, Log Logistic, Cox, Logit, and probit regression... gives extremely useful data/predictions (compared to survival analysis Kaplan Meier...)

One of the challenges of the Likert scale is bias. When the variables are in the format of positive answers, it is possible to expect a bias towards positive preferences. Then, the negation concept is often introduced with the aim of bias control. However, this can be counterintuitive, as it requires greater cognitive engagement and can lead to an increase in errors. One solution, according to Friborg Martinussen

and Rosenvinge (2006), is to use Osgood's Semantic differential (1952). The authors compare the Likert scale and the semantic differential on a sample of 334 students and conclude that the semantic differential affects the reduction of bias, i.e. there was a reduction in variance, internal consistency, and a better factor structure was obtained in the case of using the semantic differential (at the same time, there was no lowering of psychometric quality). However, the Semantic Differential cannot be considered a Likert scale because it uses bipolar antonyms (Verhagen, van den Hooff and Meents, 2015).

Uebersax (2006) elaborates on the most common mistakes in the use of the Likert scale and states the problematic nature of defining the Likert scale (e.g., when a scientist says that he used a Likert scale, it is not clear what is meant by that term). The author states that the Likert scale never refers to a separate variable but to a set of several variables whose responses are added together and give an average total result on the subject of measurement. Additionally, Uebersax (2006) postulates that a rating scale producing ordered category data (e.g., How often do you smoke cigarettes? 1. Never; 2. Once in a while; 3. 1-5 cigarettes per day; 4. More than 5 per day) is not a Likert scale. One of the ways to overcome the problem of the Likert scale is, according to Anjaria (2022), deploying the Z-number in conjunction with the Likert scale. The author introduces an algorithm for deriving Z-numbers from Likert scale data. It also explains how these concepts can be applied to convert Likert scale responses into Z-numbers.

Presentation of a new scale in social sciences

The scale consists of two levels. The first level is a binary outcome (dichotomous) where the respondent circles agree/disagree, or yes/no. The second level consists of the respondent assessing the intensity of the binary choice on a 10+10 point scale, expressed on the principle of deciles. The scale diagram is presented in Figure 2.

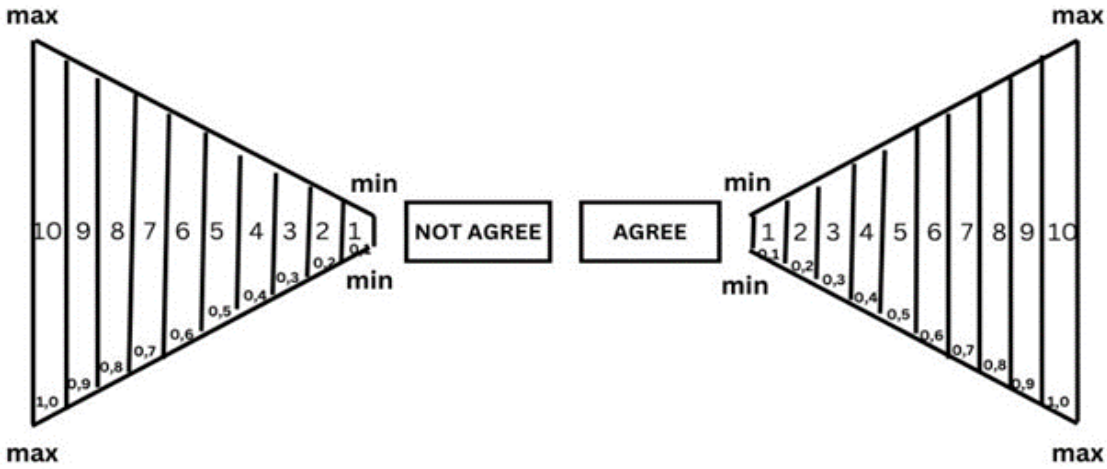


Figure 2. Binary-relative interval scale

The scale comprises a total of 20 intervals: 10 associated with the option "NOT AGREE" and 10 with the option "AGREE." Unlike traditional Likert scales, this scale does not include a neutral midpoint, as respondents do not have the option to indicate that they neither agree nor disagree.

However, a midpoint does statistically exist. For example, if the frequencies for the "NOT AGREE" and "AGREE" intervals are equal, the mean, mode, and median (MCT) would all equal 2, indicating a balanced position. In this scenario, a value of 2 suggests a state of ambivalence, reflecting both agreement and disagreement, thus representing an absence of a definitive positive or negative attitude.

The coding of the scale involves subtracting 2 for each instance marked as "Not AGREE" and adding 2 for each instance marked as "AGREE." This process can be represented as follows: for each marked

value of intensity "Not AGREE," we use x - 2, and for each marked value of intensity "AGREE," we use x + 2. The graphical representation of this coding is illustrated in Table 2.

Table 2. *Coding scale*

	Not agree (NEGATT)										Agree (POSATT)									
X Initial scale values	1	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
		2- (x)												2+x						
Coded values	1	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3

After coding, we got a rating scale from 1.0 to 3.0. The interpretation of the numerical values on the scale represents the intensity of the attitude (figure 3).

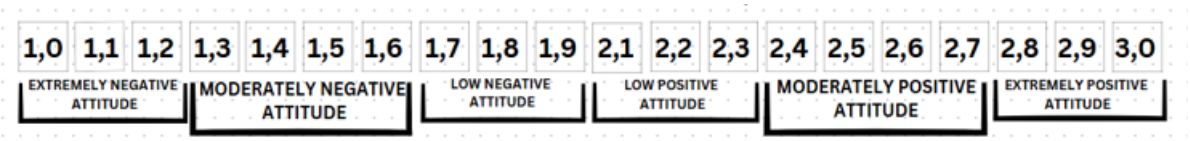


Figure 3. Intensity of the attitude

The new scale does not have an explicitly defined neutral value. However, we can determine a neutral value by calculating a measure of central tendency, such as the mean. For instance, if respondents have the same frequency across all intervals (degrees), for example, f (1)=1, f (1.1)=1, f (1.2)=1, f (1.3)=1, and so on until f (3)=1, then the mean would equal 2.00. In this scenario, the median would also be 2, and the skewness would be 0, indicating a neutral attitude. Statistically, values within the range 1,9 ≤ x ≤ 2,1 indicate a neutral midpoint, representing the absence of a specific attitude neither positive nor negative (see figure 4).

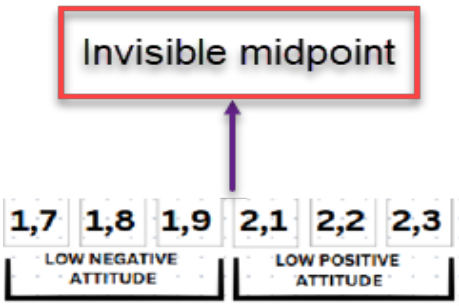


Figure 4. Invisible midpoint of scale

Advantages of the scale

In contrast to a regular Likert scale, our proposed scale has several advantages including:

1. Generating a binary outcome (discrete) in a form (yes-no, agree-not agree) suitable prior for non-parametric tests/approaches (e. g. logit and probit regression, Bayesian inference, Fisher's exact test, descriptive statistics for binary data, Log-linear…). Binary variables are cost-effective, easier to understand and interpret, and require less cognitive effort. They simplify decision-making and are less influenced by cultural differences that can arise when using a scale with multiple degrees of agreement or disagreement. Additionally, the binary choice format (such as agree/disagree or yes/no) promotes simplicity and helps streamline generalizations.
2. Existence of double measurement: binary outcomes (agree/not agree; yes/no) and additional relative interval measurement as extended measurement. The assumption is that respondents can relatively well distinguish 10 degrees (intervals), the intensity of agreement (disagreement) with a certain statement. The new scale is based on 10 relatively identical intervals that represent centiles, and it is easier for the respondent to distinguish the intensity on a scale up to 10 than, for example,

to 7 or 8. It is better and more precise to distinguish using 10 parts (deciles) because it is based on tens (10,20,30...). No, of course, this is conditional and the term relatively interval scale is used. Introduction of 10 intervals (deciles) separately for negative and positive attitude of impact on measurement precision. For example, Figure 5 shows the graphical distribution of responses to variable V1 (empirical part) using the Likert scale and the new scale. The introduction of 10 intervals (deciles) for both negative and positive attitudes improve the precision of measurement. For instance, Figure 4 illustrates the graphical distribution of responses to variable V1 (empirical part) using both the Likert scale and the new scale.

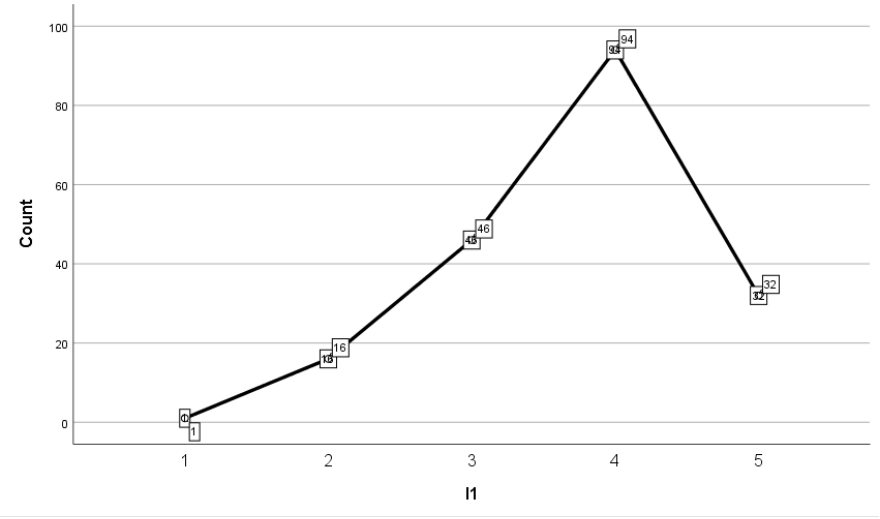


Figure 5a. Likert

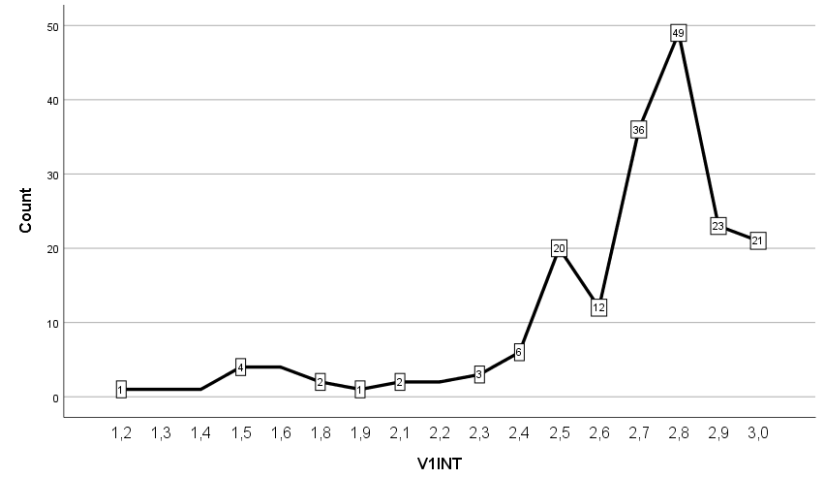


Figure 5b. New scale

Figure 5a (Likert scale responses) illustrates the distribution of responses to variable I1. The results indicate that: 1 respondent completely disagrees, 16 respondents partially disagree, 46 respondents neither agree nor disagree, 94 respondents partially agree, and 32 respondents fully agree with the statement: “In most ways, my life is close to my ideal.” The distribution of responses for the same variable on the new scale involving the same respondents, is presented in Table 3.

Table 3. *Distribution of variable V1INT (new scale)*

Value	f	INTENSITY of attitude
1,0-1,2	1	Extremely negative attitude
1,3-1,6	10	Moderately negative attitude
1,7-1,9	3	Low negative attitude
2,1-2,3	7	Low positive attitude
2,4-2,7	74	Moderately positive attitude
2,8-3,0	93	Extremely positive attitude

In this example, the improved precision of the measurement is evident. On the Likert scale, 46 subjects reported a neutral value of 3, which is not reflected in the new scale. For the variable I1 (Likert), the arithmetic mean is 3.7. On a 5-point scale, this value can be challenging to interpret regarding intensity; instead, measures such as mode or median may be more appropriate. In contrast, the new scale yields a mean of 2.6, which accurately represents a real value, id est. a *moderately positive attitude* within the statistical set. However, it's important to note that the larger range of intervals on the new scale can result in extreme values that distort the arithmetic mean. Therefore, the new scale is recommended, particularly when dealing with large sample sizes.

3. The scale values, after coding, are expressed with one decimal place, enhancing the precision of the measurement.

1	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3

The proposed scale enhances the accuracy of descriptive statistics, as the calculated arithmetic means reflect actual values on the scale, rounded to one decimal place. For instance, if the mean is 1.63 (NEGATT), the value 1.6 is a valid point on the scale, meaning the mean is a real, measurable value rather than an abstract one.

4. The ability to perform differential statistical presentation and analysis for each binary choice 'Not agree' (NEGATT for negative attitude) or 'I agree' (POSATT for positive attitude), as illustrated in Figure 6.

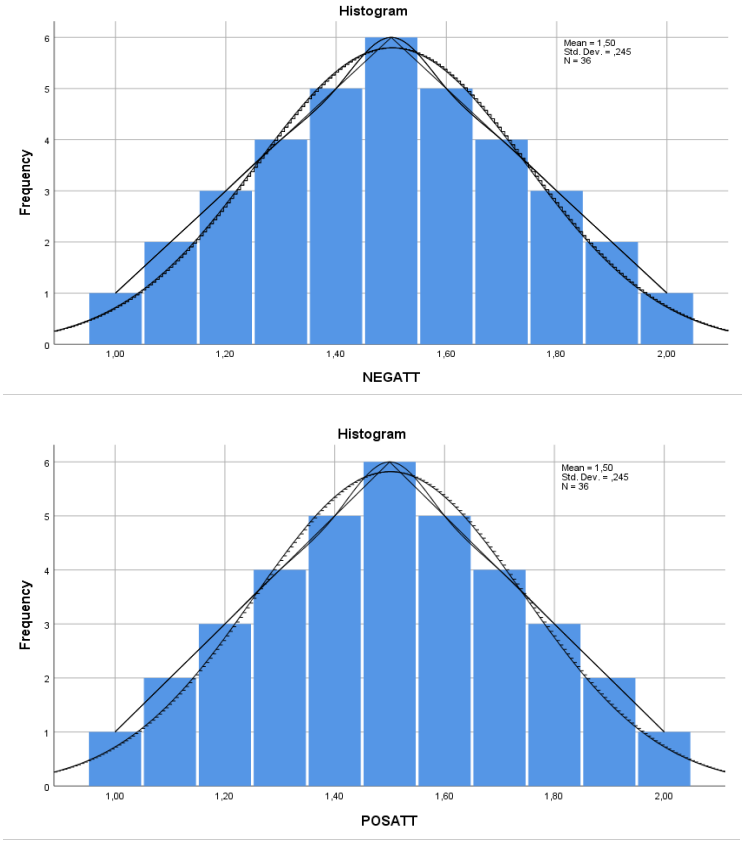


Figure 6. Distribution (normal) for NEGATT and POSATT

5. The invisible midpoint occurs between the ranges of NEGATT (0.1-1, coded 1-1.9) and POSATT (0.1-1, coded 2.1-3), with the coded value 2 representing the invisible midpoint
6. The probability of missing responses is reduced because the respondent is presented with a clear binary choice, unlike in the Likert scale. This binary format requires less cognitive effort from the respondent.

Empirical part

To gain insight into the descriptive characteristics of the new scale, a study on life satisfaction was conducted with a sample of 190 students at the University of Zagreb. The respondents completed a modified version of the Satisfaction with Life Scale (SWLS) using a 5-point Likert-type scale. They then re-evaluated their level of agreement with the same variables, this time using the new scale, which included a total of 12 variables.

The descriptive characteristics of all 12 variables on the new scale show that the coefficient of variation (CV) is lower than that of the Likert scale. This indicates better consistency and less variability, meaning that the mean is a more accurate representation of the actual values. Additionally, the standard errors of the arithmetic means are smaller, suggesting that the sample means more accurately reflect the population values. As a result, the confidence intervals are narrower due to the smaller standard error (SE), leading to greater certainty about the estimates. Similarly, a reduction in the standard error enhances the power of the statistical tests, thereby decreasing the likelihood of a Type II error. Moreover, the arithmetic mean values represent the actual values (to one decimal place), which is often not the case with the Likert scale. Although the values of skewness and kurtosis are higher, they remain within the acceptable limits according to Kline (2016).

Table 4. *Descriptive values of Likert scale and new scale*

	Range	Minimum	Maximum	Mean		Std. Deviation	Coefficient of variation	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Skewness
I1	4	1	5	3,74	,062	,858	0,23	-,495
V1INT	1,8	1,2	3,0	2,64	,0265	,3629	0,14	-2,119
I2	4	1	5	4,29	,061	,840	0,20	-1,344
V2INT	1,8	1,2	3,0	2,70	,0244	,3351	0,12	-2,019
I3	3	2	5	4,24	,057	,776	0,18	-,872
V3INT	1,7	1,3	3,0	2,74	,0218	,2988	0,11	-2,634
I4	3	2	5	4,09	,061	,836	0,20	-,559
V4INT	1,7	1,3	3,0	2,68	,0239	,3282	0,12	-2,039
I5	4	1	5	3,50	,090	1,236	0,35	-,314
V5INT	2,0	1,0	3,0	2,42	,0432	,5914	0,24	-,851
I6	4	1	5	4,07	,068	,934	0,23	-,810
V6INT	1,9	1,1	3,0	2,68	,0286	,3917	0,15	-2,019
I7	4	1	5	3,72	,073	1,003	0,27	-,567
V7INT	2,0	1,0	3,0	2,53	,0353	,4842	0,19	-1,580
I8 (recoded)	4	1	5	3,53	,090	1,231	0,35	-,417
V8INT (recoded)	2,0	1,0	3,0	2,23	,0489	,6689	0,30	-,403
I9	4	1	5	3,44	,075	1,035	0,30	-,383
V9INT	2,0	1,0	3,0	2,36	,0420	,5730	0,20	-,895
I10	3	2	5	3,98	,059	,807	0,20	-,331
V10INT	1,7	1,3	3,0	2,62	,0265	,3616	0,14	-1,683
I11	4	1	5	3,18	,073	1,008	0,32	-,180
V11INT	2,0	1,0	3,0	2,18	,0434	,5914	0,27	-,493
I12	4	1	5	2,84	,087	1,203	0,42	,357
V12INT	2,0	1,0	3,0	1,99	,0517	,7068	0,36	-,021

V1(I1) In most ways, my life is close to my ideal; V2(I2) The conditions of my life are excellent; V3(I3) I am satisfied with my life; V4(I4) So far I have gotten the important things I want in life; V5(I5) If I could live my life over, I would change almost nothing; V6(I6) When I watch others live, I feel satisfied; V7(I7) I am satisfied with my appearance; V8(I8) I would be more open to living in a different part of the world; V9(I9) I am happy with my finances;V10(I10) A bright future awaits me;V11(I11) The educators' salary will be enough for me to have a comfortable and happy life;V12(I12) I am satisfied with the position of teacher/ educator in society.

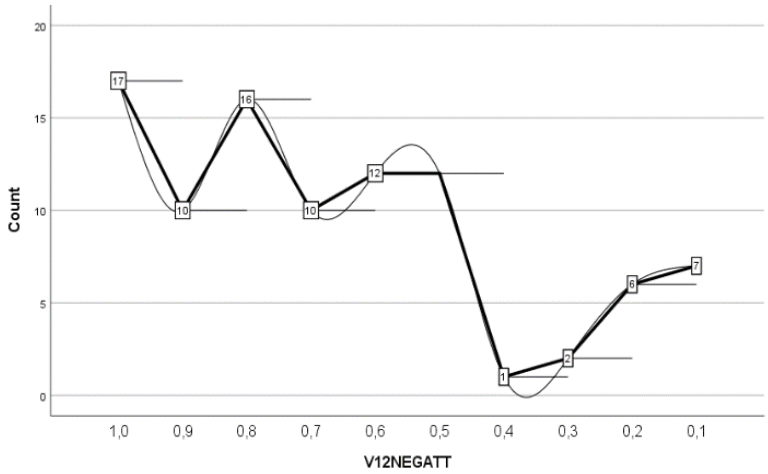
The new scale also presents more accurate, precise quantiles (quartiles, deciles) which results in better interpolation. Table 5 shows the quantile values for the Likert and new scales.

Table 5. *Descriptive values of scales*

		New scale												Likert											
		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12
Percentiles	10	2,29	2,30	2,40	2,39	1,48	2,28	1,60	1,00	1,50	2,10	1,27	1,10	3,00	3,00	3,00	3,00	2,00	3,00	2,00	1,00	2,00	3,00	2,00	1,00
	20	2,50	2,50	2,60	2,50	1,70	2,50	2,30	1,10	1,70	2,50	1,50	1,20	3,00	4,00	4,00	3,00	2,00	3,00	3,00	1,00	3,00	3,00	2,00	2,00
	25 q1	2,52	2,60	2,70	2,60	1,90	2,60	2,40	1,10	1,90	2,50	1,60	1,30	3,00	4,00	4,00	4,00	3,00	3,00	3,00	1,00	3,00	3,00	3,00	2,00
	30	2,60	2,70	2,70	2,60	2,24	2,60	2,50	1,20	2,30	2,50	1,71	1,40	3,00	4,00	4,00	4,00	3,00	4,00	3,00	2,00	3,00	4,00	3,00	2,00
	40	2,70	2,70	2,80	2,70	2,50	2,70	2,60	1,40	2,48	2,60	2,20	1,50	4,00	4,00	4,00	4,00	3,00	4,00	4,00	2,00	3,00	4,00	3,00	2,00
	50	2,70	2,80	2,80	2,80	2,60	2,80	2,70	1,50	2,60	2,70	2,35	2,10	4,00	4,00	4,00	4,00	4,00	4,00	4,00	2,00	4,00	4,00	3,00	3,00
	60	2,80	2,90	2,90	2,80	2,80	2,90	2,74	1,90	2,70	2,80	2,50	2,40	4,00	5,00	5,00	4,00	4,00	4,00	4,00	3,00	4,00	4,00	3,00	3,00
	70	2,80	2,90	2,90	2,90	2,90	2,90	2,80	2,30	2,79	2,80	2,60	2,60	4,00	5,00	5,00	5,00	4,00	5,00	4,00	3,00	4,00	4,00	4,00	3,00
	75q3	2,80	2,90	2,90	2,90	2,90	3,00	2,90	2,50	2,80	2,90	2,70	2,70	4,00	5,00	5,00	5,00	5,00	5,00	4,00	3,00	4,00	5,00	4,00	4,00
	80	2,90	3,00	2,90	2,90	3,00	3,00	2,90	2,50	2,80	2,90	2,70	2,74	4,00	5,00	5,00	5,00	5,00	5,00	5,00	4,00	4,00	5,00	4,00	4,00
	90	3,00	3,00	3,00	3,00	3,00	3,00	3,00	2,70	3,00	3,00	2,80	2,90	5,00	5,00	5,00	5,00	5,00	5,00	5,00	4,00	5,00	5,00	4,00	5,00

In terms of reliability testing, the new scale demonstrated greater reliability, with a Cronbach's alpha of 0.804, compared to α=0.787 for the Likert scale. Additionally, the Intraclass Correlation Coefficient (ICC) was also higher for the new scale (0.804) than for the Likert scale (0.787).

One of the key advantages of this scale is that it allows for the division of attitudes into positive (POSATT) and negative (NEGATT) categories. This setup facilitates a detailed analysis of the intensity of both positive and negative attitudes, rated on a scale from 1 to 10. Additionally, descriptive statistics can be calculated separately for each category and compared against one another. This approach is especially useful in economic research, where understanding the intensity of separate attitudes, both positive and negative, is critical. For instance, in our example referenced as V12INT, respondents are polarized regarding their satisfaction with the position of teachers and educators in society, with a mean score of 1.99, a minimum of 1, and a maximum of 3. This distribution is not normal. By separating respondents into POSATT and NEGATT categories, we gain better insight into the nuances of their satisfaction and dissatisfaction levels. The binary choice data also reveal that 49.7% of respondents are dissatisfied (do not agree) with the position of teachers and educators in society, while 50.3% are satisfied (agree). In Figures 7a and 7b, the variable V12INT is displayed, highlighting the separate NEGATT and POSATT groups.



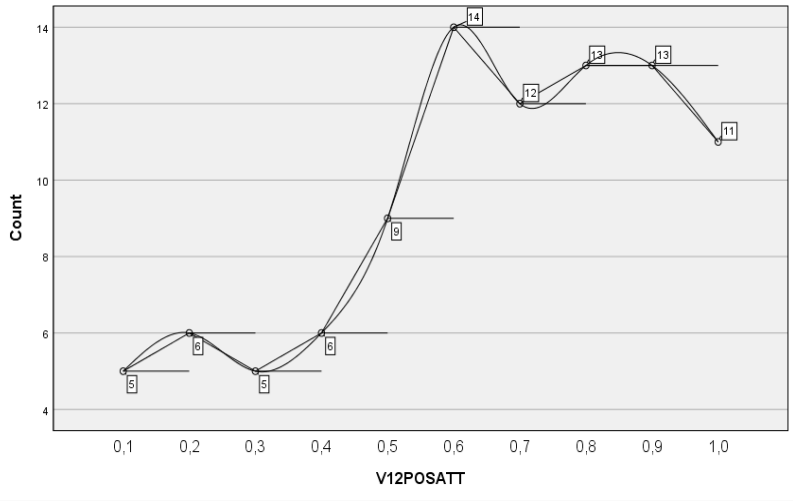


Figure 7a & 7b. NEGATT & POSSAT

Figure 7a illustrates the distribution of the NEGATT variable (V12). For instance, 16 respondents expressed a disagreement intensity of 8 out of 10 regarding the statement associated with the variable (V12), while 6 respondents indicated a disagreement intensity of only 2 out of 10 (as shown in Figure 7a). In comparison, the POSATT variable in Figure 7b shows that 14 respondents agreed with the statement at an intensity of 6 out of 10, and 11 respondents showed maximum agreement with a rating of 10 out of 10. This indicates that they are extremely satisfied with the role of teachers and educators in society (see Figure 7b).

In the case of separating respondents into POSATT and NEGATT for a certain variable(s), we can compare them separately, display descriptive characteristics, calculate the values of MCT, MV, specificities of sampling distributions, and use statistical tests to test differences and relationships between subsamples (POSATT and NEGATT). Although the sampling distribution of the variable deviates significantly from the normal distribution (attitudes are polarized), separating the distribution of the variable into NEGATT and POSATT provides the possibility of separate distributions, in this case significantly closer to the normal distribution; p-p plot (figures 8a, b, c).

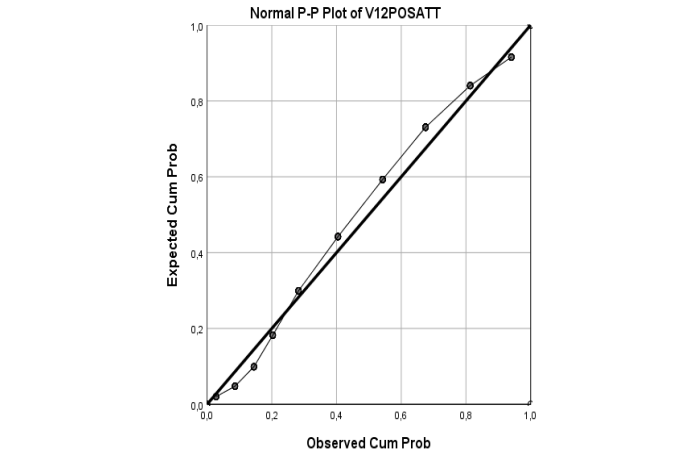


Figure 8a. P-P plot POSATT V12INT

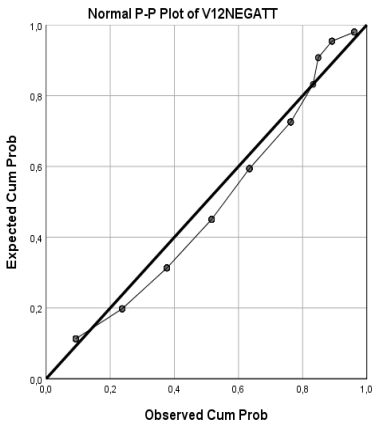


Figure 8b. P-P plot NEGATT V12INT

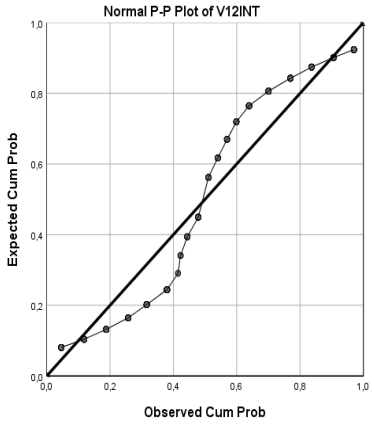


Figure 8c. P-P plot V12INT

The introduction of a new scale that includes binary choices is a valuable enhancement, as it allows for prior non-parametric analysis. In the area of predictive role testing, the new scale has the a priori option of using binary choice (as a DV) for nonlinear regression; e.g., logistic regression, Probit, Cox PHR…, and also as an additional relative interval measurement and the possibility of using linear regressions (nonlinear, too). Binary choice scales are extremely useful in a predictive sense and are often used in medicine, technical sciences. For example, binary choice V7 shows a graph of predicted probability for the variable age* V7 (figure 9).

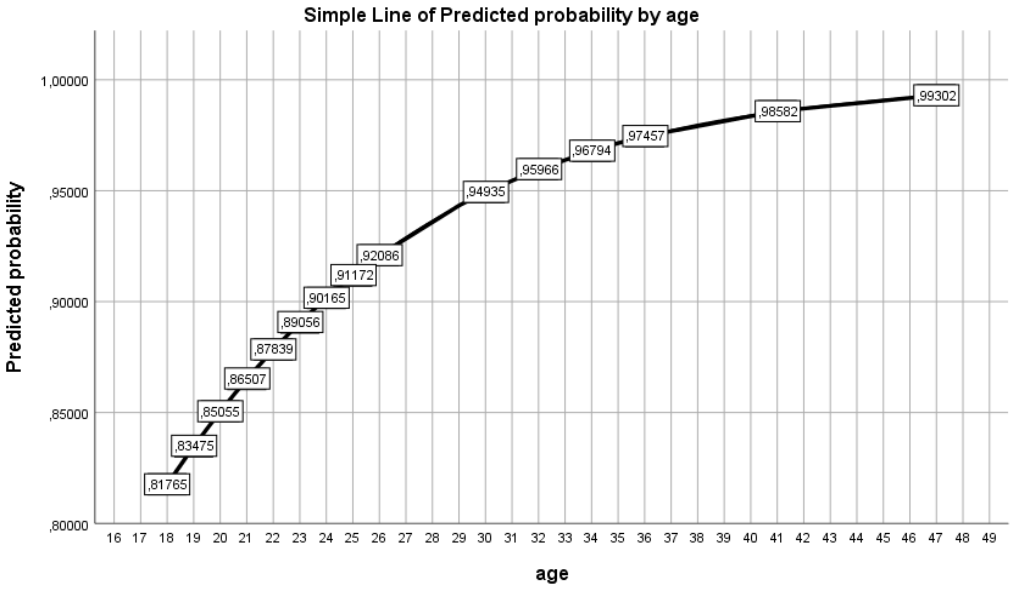


Figure 9. Predicted probability

The probability of students aged 20 in the sample being satisfied with their appearance is 83%, while for students aged 41 who are part-time students, the satisfaction rate is as high as 98%. Predicted probabilities (log-odds, odds, and probabilities) provide extremely valuable statistical data and should be utilized more in social science research. However, when using a 5-point Likert scale, the neutral value (3) complicates the ability to make a binary choice.

The new scale, which considers an interval range of 10+10, allows for the development of more complex regression models. However, this complexity can lead to issues such as overfitting or the misinterpretation of non-linear relationships between variables. Despite these potential challenges, this scale yields more accurate predictive results than the Likert scale. For instance, the interpretations of beta values are more precise, and also standardized regression residuals (see Figure 10a, b), which can also influence other regression parameters.

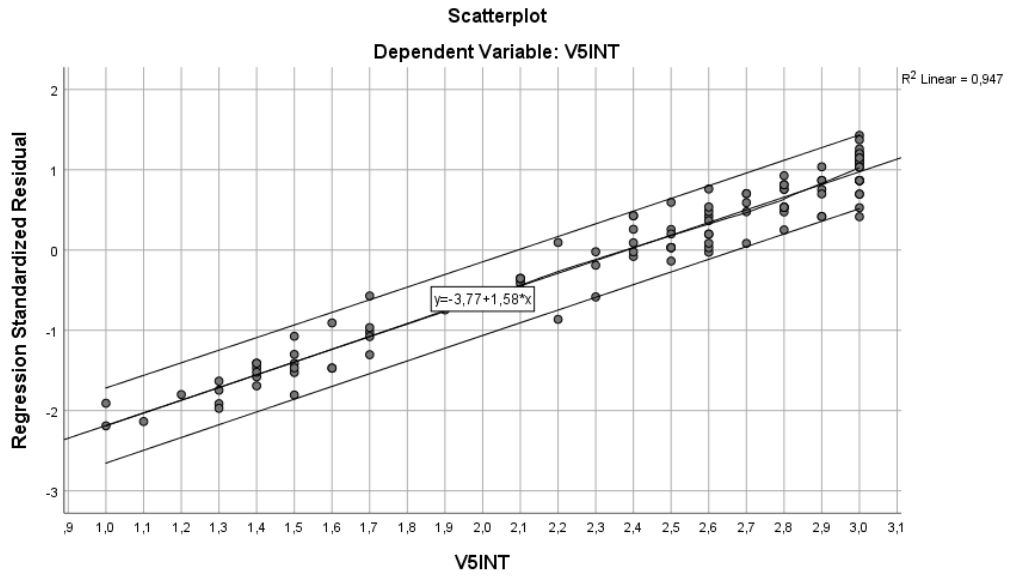


*Figure 10a. RSR V5*academic grade (New scale)*

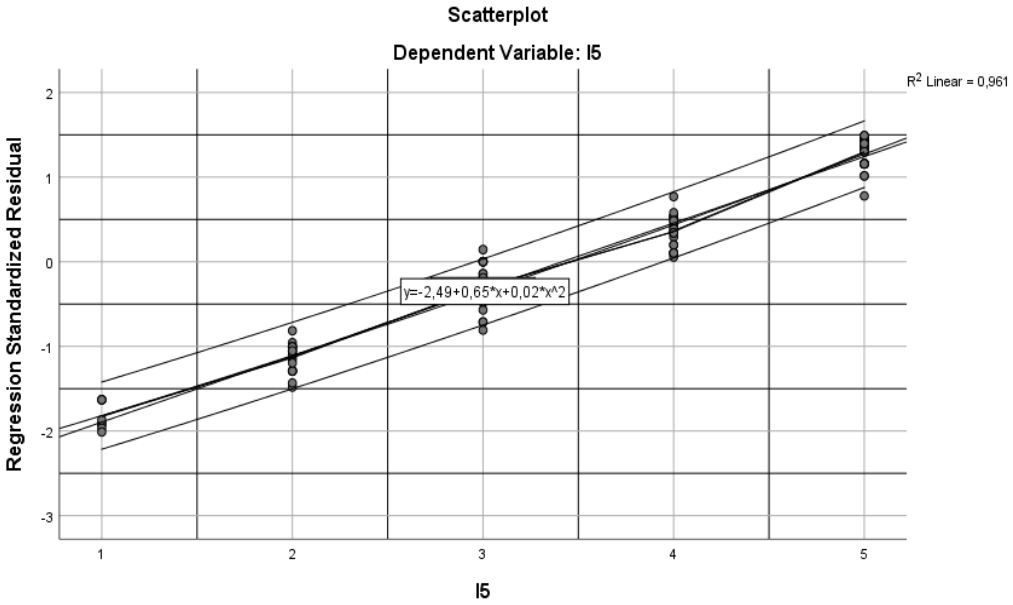


*Figure 10b. RSR V5*academic grade (Likert)*

Conclusion

The new scale offers several advantages over the traditional Likert scale. It incorporates a binary choice, which is particularly useful for predictive modeling, along with a relative interval scale featuring 10+10 intervals. In social sciences, especially in attitude research, it is not expected that absolute distinctions can be made when assessing the strength of agreement or disagreement. However, respondents are expected to recognize differences across 10 distinct intervals (NEGATT-POSATT; min-max).

The tolerance for distinction error is +-1 scale interval, i.e., 5% between intervals on a 10+10 scale. Since the respondent evaluates only one dimension at a given time (agreement or disagreement), but we use a total scale of 20 intervals, it is reasonable to expect some error in the precision of distinguishing adjacent intervals.

The scale consists of 10 degrees of agreement (POSATT) and 10 degrees of disagreement (NEGATT). After coding, the scale features 20 rank intervals or degrees (1, 1.1, 1.2, ..., up to 3). This design allows for high resolution in measurement. Respondents are expected to distinguish between more

distant points (for example, 8 vs. 10) more accurately than between neighboring intervals (such as 8 vs. 9). Due to the limited ability to differentiate within the 10 degrees, we assume an error of ±1 degree on this scale. When considering the entire 20-degree scale, this error is proportionately adjusted to 5% (1 out of 20). If a respondent estimates the intensity of negative attitudes between 8 and 10, this is likely a more precise estimate. However, if the choice is between 8 and 9, the potential for error increases because neighboring values are harder to distinguish. This error can impact the fine-tuning of statistical analyses, such as in reliability calculations or factor analysis. Therefore, the precision is 95%, or there is a 5% probability that subjects will misjudge the intensity due to difficulty distinguishing between adjacent intervals (Figure 11).

This paper highlights various benefits of the new scale compared to the Likert scale, both in descriptive and inferential statistics. The new scale significantly enhances statistical analysis capabilities by separating the data set into positive (POSATT) and negative (NEGATT) attitudes. This allows for differential analyses of positive and negative attitudes and improves prediction modeling. Such capabilities are particularly valuable in economic research, as they enable separate analyses and comparisons of positive and negative preferences (NEGATT and POSATT). The dual measurement offered by the scale, including binary outcomes (agree/disagree; yes/no) along with additional relative interval measurements, expands the range of statistical analyses available. This can lead to more accurate, objective, and reliable conclusions.

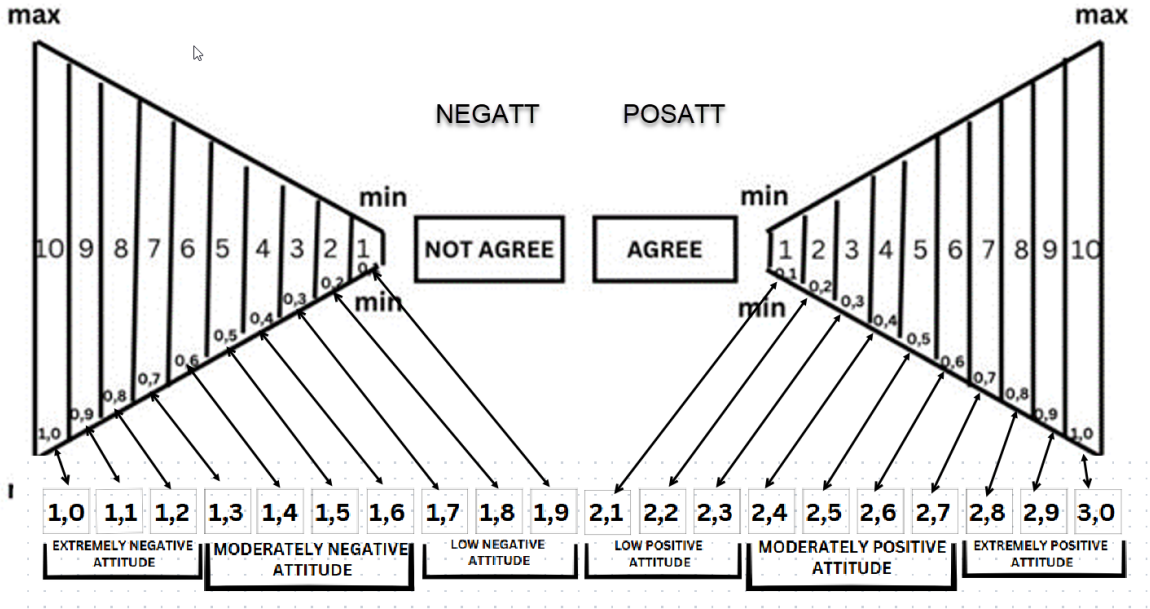


Figure 11. New scale with coding

It is recommended to use the scale for categories with large sample sizes. Further research on the new scale is anticipated, taking into account its distinct advantages as well as certain limitations.

Limitations of the new scale

The Likert scale has been the subject of criticism in many scientific journals for decades. Despite this, it remains dominant in a number of scientific disciplines. The new scale is statistically more justified, more precise, and more accurate, although there is a need for further research on its advantages and disadvantages.

Acknowledgements

I would like to thank The University of Sydney, Centre for Educational Measurement and Assessment (CEMA) for fulfilling support. Special thanks to the editor and the anonymous reviewers for their constructive suggestions for improvement.

Conflict of Interest

Author have no conflict.

Statement of Disclaimer

Funded by the European Union - NextGenerationEU. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the European Commission can be held responsible for them

Generative AI Statement

During the preparation of this article, we used ChatGPT for translation from Croatian to English language and for additional grammar and clarity improvement. The author reviewed and edited all content as needed and takes full responsibility for the published version.

References

Adams, E.W., Fagot, R.F., Robinson, R.E. (1965). A Theory of Appropriate Statistics. *Psychometrika*, 30(2), 99-127.

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

Anjaria, K. (2022). Knowledge Derivation from Likert Scale Using Z-numbers. *Information Sciences,* 590, 234-252. https://doi.org/10.1016/j.ins.2022.01.024

Aybek, Eren C., Toraman, C. (2022). How Many Response Categories are Sufficient for Likert-type Scales? An Empirical Study Based on the Item Response Theory. *International Journal of Assessment Tools in Education,* 9(2), 534-547. https://doi.org/10.21449/ijate.1132931

Bishop, Phillip A., Herron, Robert L. (2015). Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *International Journal of Exercise Science,* 8 (3), 297-302. PMID: 27182418; PMCID: PMC4833473. https://doi.org/10.70252/LANZ1453

Boote, A. S. (1981). Reliability Testing of Psychographic Scales: Five-Point or Seven-Point? Anchored or labeled? *Journal of Advertising Research,* 21(5), 53-60.

Brown, G., Widing, R. E., II, & Coulter, R. L. (1991). Customer Evaluation of Retail Salespeople Utilizing the SOCO Scale: A Replication, Extension, and Application. *Journal of the Academy of Marketing Science,* 19, 347-351.

Carifio, J., Perla, R.J. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths, and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences* 3(3), 106-116. https://doi.org/10.3844/jssp.2007.106.116

Chang L. (1994). A Psychometric Evaluation of 4-Point and 6-Point Likert-Type Scale in Relation to Reliability and Validity. *Applied Psychological Measurement,* 18(3), 205-215. https://doi.org/10.1177/014662169401800302

Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement,* 56(10), 15-23. https://doi.org/10.1002/pfi.21727

Cox, E.P. (1980). The Optimal Number of Response Alternate for a Scale: A review. *Journal of Marketing Research,* 17(4), 407-422. https://doi.org/10.1177/002224378001700401

De Vellis, R.F. (2003). Scale Development, Theory and Applications (2nd ed., Vol. 26). Thousand Oaks, CA: Sage Publications.

Finstad, K. (2010). The Usability Metric for User Experience. Interactions, 17(1), 11–15.

Friborg, O., Martinussen, M., Rosenvinge, Jan H. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences*, 40(5), 873-884. https://doi.org/10.1016/j.paid.2005.08.015

Garner, W.R. (1960). Rating scales, discriminability and information transmission. *Psychological Review,* 67(6), 343-352.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. Review of Educational Research, 42(3), 237-288. https://doi.org/10.3102/00346543042003237

Green, P.E., Rao, V.R. (1970). Rating Scales and Information Recovery: How Many Scales and Response Categories to Use? *Journal of Marketing,* 34(3), 33-39.

Guerra, Andrea L., Gidel, T., Vezzetti, E. (2016). Toward a common procedure using Likert and Likert-type scales in small groups comparative design observations. DS 84: Proceedings of the DESIGN 2016 14th International Design Conference, Dubrovnik, Croatia.

Harris, R.J. (1975). A Primer of Multivariate Statistics. New York: Academic Press.

Hernández, A., Muñiz, J., & García-Cueto, E. (2000). Comportamiento del modelo de respuesta graduada en función del número de categorías de la escala [Graded model trend based on the number of the response categories of the scale]. *Psicothema,* 12(2), 288-291.

Jakobsson, U. (2004). Statistical presentation and analysis of ordinal data in nursing research. *Scandinavian Journal of Caring Sciences*, 18(4), 341-445.

Jamieson, S. (2000). Likert Scales: How to (ab)use Them. *Medical Education,* 38(12), 1212-1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Kline, Rex B. (2016). Principles and practice of structural equation modeling (4th ed.). Guilford Press.

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

Leung, Shing-On (2011). A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales. *Journal of Social Service Research,* 37(4), 412-421. https://doi.org/10.1080/01488376.2011.580697

Lewis, J.P., Traill, A. (1999). Statistics Explained. Harlow, England: Addison-Wesley pp. 377.

Lozano, Luis M., García-Cueto, E., Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology European Journal of Research Methods for the Behavioral and Social Sciences*, 4(2),73-79. https://doi.org/10.1027/1614-2241.4.2.73

McCallum, D. M., Keith, B. R., & Wiebe, D. J. (1988). Comparison of response formats for Multidimensional Health Locus of Control Scales: Six levels versus two levels. Journal of Personality Assessment, 52(4), 732-736. https://doi.org/10.1207/s15327752jpa5204_12

Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 101(2), 343-352. https://doi.org/10.1037/0033-295X.101.2.343

Moreno, R., Martínez, R.J., & Muñiz, J. (2004). Directrices para la construcción de ítems de elección múltiple [A guide to the multiple choice items development]. *Psicothema,* 16, 490-497.

Muñiz, J., García-Cueto, E., & Lozano, L. M. (2005). Item Format and the Psychometric Properties of the Eysenck Personality Questionnaire. *Personality and Individual Differences,* 38, 61–69. https://doi.org/10.1016/j.paid.2004.03.021

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric Theory (3rd edition). McGraw-Hill.

Osgood, C. E. (1952). The Nature and Measurement of Meaning. *Psychological Bulletin,* 49, 226-233.

Pearson, B., Lacombe, D., Khatun, N. (2024). Likert Scale Variables in Personal Finance Research: The Neutral Category Problem. *Econometrics* 12(4),33. https://doi.org/10.3390/econometrics12040033

Pett, M.A. (1997). Nonparametric Statistics for Health Care Research. London: SAGE Publications.

Preston, Carolyn C., Colman, Andrew M. (2000). Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences. *Acta Psychologica* 104(1), 1-15. https://doi.org/10.1016/S0001-6918(99)00050-5

Schrum, M.L., Johnson, M., Ghuy, M., Gombolay, Matthew C. (2020). Four Years in Review: Statistical Practices of Likert Scales in Human-Robot Interaction Studies. HRI '20: ACM/IEEE International Conference on Human-Robot Interaction Cambridge United Kingdom, 43-52. https://doi.org/10.1145/3371382.3380739

Selimović, H., Opić, S., Selimović, N. (2023). Metodologija istraživanja u odgojno - obrazovnim znanostima. Travnik - Univerzitet u Travniku, Edukacijski fakultet.

Tanujaya, B., Prahmana, Rully Charitas I., Mumu, J. (2022). Likert Scale in Social Sciences Research: Problems and Difficulties. *FWU Journal of Social Sciences* 16(4), 89-101. https://sbbwu.edu.pk/journal/Winter2022,Vol.16,No.4/7.%20Likert%20Scale%20in%20Social%20Sciences%20Research.pdf

Uebersax J.S. (2006). Likert scales: Dispelling the Confusion. Statistical Methods for Rater Agreement website. Available at: http://john-uebersax.com/stat/likert.htm. Accessed: 12/12/2024.

Verhagen, T., van den Hooff, B., Meents, S. (2015). Toward a Better use of the Semantic Differential in IS Research: An Integrative Framework of Suggested Action. *Journal of the Association of Information Systems* 16 (2), 108-143.

Yaska, M., & Nuhu, B. M. (2024). Assessment of Measures of Central Tendency and Dispersion Using Likert-Type Scale. *African Journal of Advances in Science and Technology Research*, 16(2), 108-143. https://doi.org/10.62154/ajastr.2024.016.010379