The Sankalpa: International Journal of Management Decisions
ISSN: 2454-7425
Volume 12, Issue 1 (January - June 2026)

The Most Important Statistical Methods Used With The Likert Cumulative Scale in the Human and Social Sciences

Dr HELIS Djenidi
Université Abou Bekr Belkaid – Tlemcen Algeria
Djenidi.helis@univ-tlemcen.dz
helisdjenidi@gmail.com
ORCID: https://orcid.org/0009-0003-3179-5504

Submission: 20.08.2025; Acceptance: 01.01.2026 ; Publication: 03.02.2026

Abstract:

The Likert cumulative scale is a cornerstone of quantitative research in human and social sciences, enabling the measurement of attitudes, perceptions, and behaviors through structured response formats. This abstract outlines the most important statistical methods employed to analyze Likert-type data: Descriptive Statistics: Frequencies, percentages, means, and standard deviations are used to summarize response distributions and central tendencies. Reliability Analysis: Cronbach's alpha is the standard tool for assessing internal consistency across Likert items, ensuring scale reliability. These methods collectively enhance the rigor and interpretability of Likert-based research, supporting robust conclusions in fields such as psychology, education, sociology, and political science.

Keywords Likert cumulative scale, human and social sciences,

Introduction:

The Likert cumulative rating scale is a behavioral measurement method used in psychological testing, developed by psychologist Rensis Likert. It is commonly used in surveys, especially in the field of statistics, and relies on responses that indicate the degree of agreement or disagreement with a given statement.

In disciplines such as psychology, sociology, education, and anthropology, the Likert scale facilitates the exploration of complex human experiences by transforming qualitative insights into measurable variables. Its simplicity, flexibility, and statistical compatibility make it a cornerstone in survey research and empirical studies, helping scholars understand how individuals think, feel, and behave in various contexts.

In our study on job satisfaction and performance, we reviewed the most effective methods commonly used in behavioral measurement, with a particular focus on the Likert scale. As one of the most widely adopted tools in social and human sciences, the Likert scale enables researchers to capture the intensity of respondents' attitudes, perceptions, and experiences. By presenting statements related to workplace dynamics and asking participants to indicate their level of agreement, we were able to quantify subjective responses and analyze patterns that reflect employee satisfaction and performance outcomes. This approach provided a reliable framework for assessing psychological and organizational variables in a structured and statistically meaningful way.

1. **Likert's Cumulative Rating Method:**

In 1932, Likert published a paper titled *A Technique for the Measurement of Attitudes* in the *Archives of Psychology*. He proposed a new method based on the equal interval scale. This method involves measuring attitudes toward a subject by presenting statements that reflect the attitude being studied. Each statement is followed by five response options, and the individual is asked to indicate their level of agreement with each statement by selecting one of several alternatives.

1.1 Definition of the Scale

The Likert scale is the sum of responses collected across several indicators, each consisting of two parts:

- **Stem**: A sentence or question that identifies a specific behavior.
- **Scale**: A measurement tool used to determine the degree of agreement or disagreement with the stem.

The Likert scale assumes that the intensity (strength) of an attitude is always linear i.e., on a continuum from strongly agree to strongly disagree, and makes the assumption that attitudes can be measured (Likert, 1932). Likert scales with odd response choice categories (5, 7, 9, and 11) are generally concentrated in the middle of the scale and lead to more items whose weights are assigned mostly in the middle of verbal descriptions. (Kusmaryono, I., Wijayanti, D., & Maharani, H. R 2022.p626)

Example: Stem: Are you in favor of birth control?

Scale:

- Strongly Agree □
- Agree □
- Not Sure □
- Disagree □
- Strongly Disagree □

Other formats include:

- **Three-point scale**: Yes □ Not Sure □ No □
- **Seven-point scale**:
  (Strongly Agree – Agree – Somewhat Agree – Not Sure – Somewhat Disagree – Disagree – Strongly Disagree)

1.2 Steps for Using the Likert Scale

1. Construct a scale with selected statements, some phrased positively and others negatively, each followed by five response options ranging from strong agreement to strong disagreement.
2. Select a sample for the study and collect responses.
3. Sum the scores for all statements and individuals in the sample.
4. Examine the consistency of responses for each statement and compare it with the total score.
5. Eliminate statements that do not align with the rest.
6. Finalize the scale statements and recalculate the total scores after removing inconsistent items.

Researchers should be familiar with item analysis techniques, reliability and validity testing, and general measurement principles (Achoui, 1992, p.193).

1.3 Specific Problems of the Summated Rating Scale (Likert Scale)

The questionnaire is a key tool for data collection, and the Likert scale is a statistical method that helps researchers identify respondents' attitudes. Its advantages include ease of aggregation and classification into dimensions, allowing variables to be measured according to the study's concepts and themes. That's why most social research focused on attitudes relies on this scale.

However, researchers must be aware of its limitations to use it correctly. For example, some use a four-point scale and omit the "Neutral" option, which is a core component of the scale. Replacing it with "Somewhat Agree" introduces bias (3 agree vs. 2 disagree options).

Another issue is the lack of attention to the phrasing of indicators—positive or negative. All questions within a theme should be consistently phrased, with negations included to avoid misleading the mean score. Statistical software cannot detect this; researchers must ensure consistency when formulating hypotheses and defining concepts.

According to David R. Hodge and David Gillespie (2003, pp. 45–55), the Likert scale may not fulfill its intended purpose due to cognitive complexity. Respondents often struggle to choose among five options, especially the middle one, which is used to avoid answering. They suggest more precise response keys, such as two statements connected by a 10-point scale from 0 to 10:

Example: Does my religious belief affect me?

Absolutely, in all aspects of my life	10	9	8	7	6	5	4	3	2	1	Not at all

Personal Insight

Based on my own research experience, I believe the Likert scale should be used partially, representing only one dimension or variable within a broader measurement framework. This allows for more comprehensive statistical analysis and opens the door to using more complex questions.

1.4 Data Table Analysis

Each response level is assigned a numerical value:

Response	Value
Strongly Disagree	01
Disagree	02
Neutral	03
Agree	04
Strongly Agree	05

Source: Prepared by the researcher.

To calculate the range for each value, apply the formula:

Range = $\frac{(N-1)}{N}$

Where **N** is the number of Likert scale points.

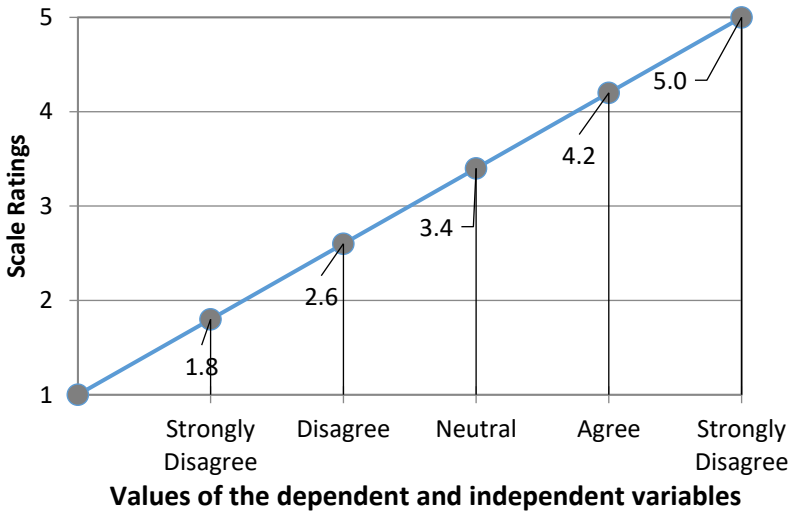


Chart N°01 presents the score scale of the dependent and independent variables.

1.5 Calculating Averages in the Likert Cumulative Scale

To calculate the average for each indicator, multiply the assigned value by the number of responses (frequencies), then divide the total by the number of individuals in the research population.

To calculate the average for each dimension, sum the averages of the indicators and divide by the number of indicators.

To calculate the overall average of the scale, sum the averages of all dimensions and divide by the number of dimensions.

1.6 The Scale Structure

A questionnaire is designed to collect the necessary data and information, covering the following main sections:

- **Section 1**: Personal data of the respondent
- **Section 2**: Questions related to the independent variable, with its dimensions defined
- **Section 3**: Questions related to the dependent variable, with its dimensions defined

Once it is clear that under which rules the items are categorized and what the direction of inquiry is, it becomes obvious that the further statistical treatment as per their assignment into ordinal or interval scale (Sullivan, G. M., & Artino, A. R. 2013 .p 401)

2. Statistical Methods Used with the Likert Scale

- Cronbach's Alpha test to measure reliability
- Arithmetic mean and standard deviation
- Pearson correlation coefficient to examine relationships between variables
- Homogeneity test
- Normal distribution test

- Analysis of variance (ANOVA) to examine differences between means
- Multiple comparison test (Post Hoc)

2.1 Reliability of the Likert Cumulative Scale

Cronbach's Alpha is the reliability coefficient that measures the internal consistency of a scale composed of multiple indicators. This process helps reduce initial items through iterative analysis, retaining those that improve the Alpha value. Internal consistency refers to the strength of the relationship between scores of each domain and the total questionnaire score. The Alpha value ranges from 0 to 1, and the closer it is to 1, the stronger the reliability. Items that reduce the Alpha value should be removed, while those that increase it should be retained. Cronbach's Alpha analysis prevents researchers from advancing in their study with repeated errors due to measurement issues in the questionnaire.

Acceptance levels of Alpha vary depending on the study's purpose:

- For exploratory studies: Alpha > 0.7 is acceptable
- For in-depth studies: Alpha > 0.8 is required
- For group comparisons: Alpha = 0.8 is satisfactory

De Velles **(2003)** proposes the following classification (Manu & Fanny, 2009, p.53):

Reliability Level	Alpha Value
Insufficient	< 0.6
Weak	0.6 – 0.65
Acceptable	0.65 – 0.7
Good	0.7 – 0.8
Very Good	0.8 – 0.9
Consider Reducing Items	> 0.9

2.2 Arithmetic Mean of the Likert Scale

The arithmetic mean is the most common and central tendency measure. It requires a simple explanation: **Mean ($\bar{x}$) = Sum of values / Number of values** $(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} xi$ (Howell, 2008, p.35)

Example:

Table (01): Shows the Means and Standard Deviations According to Job Satisfaction Level for the Direct Supervision Axis

No.	Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Mean	Std. Dev.
		%	%	%	%	%		
1		9	17	31	144	47	3.82	0.941

	My supervisor inquires about my work, needs, and guides me during work	3,6	6,9	12,5	58,1	19		
2	I am not bothered by my supervisor's monitoring style	38	51	35	100	24	3.08	1.268
		15,3	20,6	14,1	40,3	9,7		
3	My supervisor appreciates the effort I put into work	23	17	38	115	55	3.65	1.170
		9,3	6,9	15,3	46,4	22,2		
4	I feel my supervisor treats all employees equally	18	23	49	99	59	3.64	1.155
		7,3	9,3	19,8	39,9	23,8		
5	My supervisor allows me to participate in decision-making	20	36	48	115	29	3.39	1.119
		8,1	14,5	19,4	46,4	11,7		
6	My supervisor has high knowledge and expertise in the field	18	18	41	105	66	3.74	1.145
		7,3	7,3	16,5	42,3	26,6		
7	I have a good relationship with my supervisor	10	13	29	132	64	3.92	0.972
		4	5,2	11,7	53,2	25,8		
8	I do not consider changing my unit due to issues with my supervisor	26	30	24	108	60	3.59	1.266
		10,5	12,1	9,7	43,5	24,2		
							3.6033	0.76872

(Source: Helis, 2014, p.100)

The mean values for each indicator and dimension are plotted on a Chart N°01 to assess the average acceptance or rejection levels among respondents.

2.3 Correlation Between Dependent and Independent Variables

This study aims to identify and measure the correlation between the dependent and independent variables, both measured using the Likert scale.

Pearson's linear correlation coefficient is widely used in social and human sciences. It is applied when the researcher is interested in the existence of a relationship between variables X and Y, without necessarily exploring causality. In such cases, the correlation coefficient can be calculated without performing regression analysis. (Arab Planning Institute, 2005, Issue 24)

$$R = \frac{n\sum xiyi - \sum xi \sum yi}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

Correlation vs. Regression:

- Regression explains changes in Y based on X (causal relationship)
- Correlation examines whether a relationship exists (Maureen, 2005, p.42)

Pearson's correlation and regression are suitable for categorical and ratio variables. However, many researchers apply them to ordinal variables as well (FOX, 1999, p.273).

Example:

Table (02): Pearson correlation between job satisfaction and performance

Correlation

Performance	Job Satisfaction		
,441**	1	Pearson Correlation	
,000		Sig. (two-tailed)	Job Satisfaction
248	248	N	
1	,441**	Pearson Correlation	
	,000	Sig. (two-tailed)	Performance
248	248	N	

**. The correlation is significant at the 0.01 level (two-tailed)

(Source: Helis, 2014, p.117)

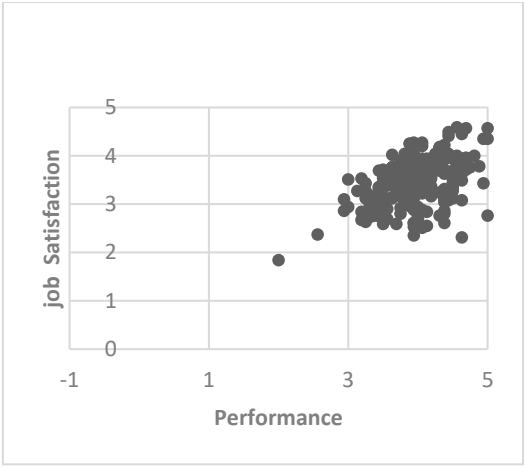
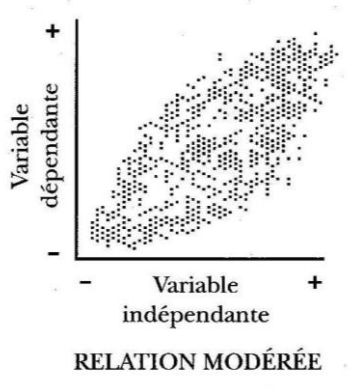


chart 02 presents the distribution of Pearson correlation coefficients between the Job Satisfaction Scale and the Performance Scale, as reported by: (Helis, 2014, p117).

There is no strict rule that defines strong or moderate correlations. However, general guidelines can help in interpretation (FOX, 1999, p.269).

	Negative Relationship					No Relationship			Positive Relationship		
r=	-1.00	-0.8	-0.6	-0.4	-0.2	**0.00**	0.2	0.4	0.6	0.8	1.00
	Perfect	Strong	Moderate	Weak		**No Relationship**	Weak	Moderate	Strong		Perfect

performance.



RELATION MODÉRÉE

Figure 01: Scatterplot of the Density Relationship (Fox, 1999, p. 269)

According to Howell (2009, p. 281), the Pearson correlation coefficient may be substantially affected by certain sample characteristics, notably the restriction of range (or variance) in either variable X or Y, and the inclusion of heterogeneous subsamples.

2.4 ANOVA Test between Dependent Variable (Likert Scale) and Personal Characteristics

One-way ANOVA (Analysis of Variance) is used to interpret social phenomena and test the null hypothesis by comparing means of the independent variable across personal characteristics.

ANOVA is widely used in psychological research and, unlike the T-test, can handle multiple group means (2, 3, 5, or more). It allows simultaneous analysis of two or more independent variables (Howell, 2008, p.306).

ANOVA breaks down the total variance of the dependent variable into:

- Within-group variance (due to the independent variable)
- Between-group variance (among independent variable categories) (FOX, 1999, p.233)

2.4.1 Conditions for Using ANOVA:

- Dependent variable must be quantitative (ordinal sometimes accepted)
- Random sampling
- Independence between group means
- Normal distribution of the dependent variable
- Homogeneity of variance across groups

Violating these conditions can affect hypothesis testing results. Minor violations may be tolerated, but major ones compromise validity (FOX, 1999, p.233).

a. Normal Distribution Test

The Kolmogorov-Smirnov test is a well-known method for testing normality, available in SPSS under non-parametric tests. However, many experts advise against using it:

- Small samples often pass the test even when not normally distributed
- Large samples often fail the test due to minor deviations (D'Agostino & Stephens, 1986): "The Kolmogorov-Smirnov test is a historical relic. It should never be used."

This test is mentioned only because it appears in some sources and researchers should be aware of its limitations (Howell, 2009, p.79).

In many cases, it's difficult to visually assess normality. Even seemingly normal distributions may not be so. A better approach is using the **Q-Q plot** (Howell, 2009, p.77).

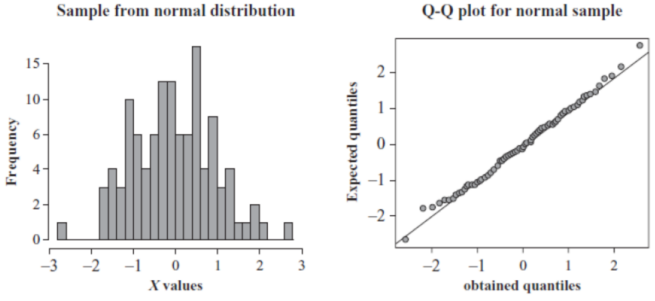


Figure 02: normal distributions (Howell, 2009, p.77).

Educators and researchers also commonly create several Likert-type items, group them into a ‘‘survey scale,’’ and then calculate a total score or mean score for the scale items. Often this practice is recommended, particularly when researchers are attempting to measure less concrete concepts, such as trainee motivation, patient satisfaction, and physician confidence—where a single survey item is unlikely to be capable of fully capturing the concept being assessed.5 In these cases, experts suggest using the Cronbach alpha or Kappa test or factor analysis technique to provide evidence that the components of the scale are sufficiently intercorrelated and that the grouped items measure the underlying variable. (Sullivan, G. M., & Artino, A. R. 2013.p542)

b. Levene’s Test for Homogeneity of Variances

As previously mentioned, one of the key conditions for using analysis of variance (ANOVA) is testing the homogeneity of variances using **Levene’s Test**.

Levene’s Test is essentially a **T-test** for the absolute or squared differences of observations from the group means or medians. If the variance of one group is larger than another, the differences from the mean or median will generally be greater in that group. Therefore, the T-test for absolute differences serves as a test for group variance. *(Howell, 2008, p.326)*

Example:

Table (03): One-way ANOVA for Educational Level and Job Satisfaction

Job Satisfaction **ANOVA à 1 facteur**

Significatio n	F	Moyenne des carrés	ddl	Somme des carrés	
,038	3,320	,739	2	1,477	Inter-groupes
		,222	245	54,498	Intra-groupes
			247	55,975	Total

Source: Helis, 2014, p.125

The **P-value (0.038)** associated with the F-statistic (F = 3.320) is statistically significant at the 0.05 level (α = 0.05). Therefore, we reject the null hypothesis that there are no statistically significant differences between the means of job satisfaction across educational levels. This suggests that **educational level affects job satisfaction** among employees of the Civil Protection Directorate in Djelfa Province.

ANOVA measures overall differences between means, not specific group differences. Statistical significance may result from a large difference between just two means among several. For this reason, statisticians developed **Post Hoc Multiple Comparison Tests** to address such cases. *(FOX, 1999, p.247)*

At this stage, we do not know which educational levels (middle school, secondary, higher education) differ in satisfaction. Therefore, we apply a **Post Hoc Test**.

“If there are five means, all differences are tested as if they were five separate steps, controlling the error rate for each null hypothesis. Tukey’s HSD is a preferred pairwise test for many due to its error control.” *(Howell, 2009, p.392)*

If group sizes are unequal or variances are not homogeneous, the same general procedure applies. Use **Tukey** for comparing all groups, and **Duncan** when comparing one group against all others. *(Howell, 2009, p.396)*

In this case, we choose **Duncan's Test** because it compares different groups with a control group.

Example:

Table (04): Duncan's Multiple Comparison Test for Educational Level and Job Satisfaction

Duncan — **Job Satisfaction**

Sous-ensemble pour alpha = 0.05		N	**Educational Level**
2	1		
	3,3627	60	Higher Education
	3,4460	160	Secondary
3,6408		28	Middle School
1,000	,373		Signification

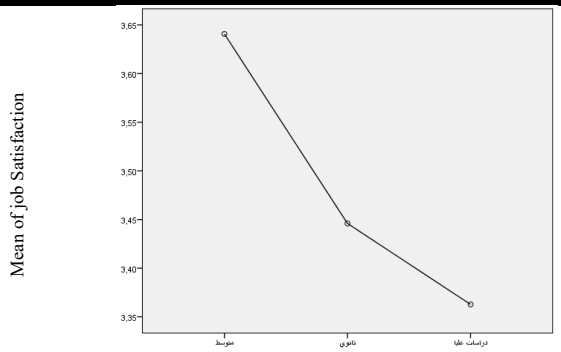


Chart (02): Distribution ( ) g to Likert Scale

Source: Hills, 2014, p.120

Duncan's Test shows that the **middle school level** has a significantly higher mean than the other two levels. In other words, employees with a middle school education are **more satisfied** than those with secondary or higher education.

2.5 Two-Sample Mean Difference Test (T-Test)

Key conditions for using the **T-test**:

- The dependent variable must be quantitative.
- The dependent variable should be normally distributed, especially in small samples (< 50).
- Variances of the dependent variable should be equal across groups homogeneity of variance or homoscedasticity . *(FOX, 1999, p.220)*

Before conducting any statistical test before data collection, we must define the test's features, including the null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

- The null hypothesis is two-tailed: we reject H_0 if $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$.
- We use a **two-tailed T-test** to compare means.

The T-value is calculated using the following formulas: *(Howell, 2009, p.208)*

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S\ \bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Example:

Table (05): Levene's Test for Marital Status and Job Satisfaction

Group statistics

Std. Error	Std. Dev.	Mean	N	Marital Status	
,03348	,44417	3,4114	176	Married	Job Satisfaction
,06375	,53337	3,5566	70	Single	

Source: Helis, 2014, p.130

Table (06): Independent Samples T-Test for Marital Status and Job Satisfaction

Test Independent samples

t-test for equality of means						Levene's test for equality of variances		Job Satisfaction
Confidence interval 95% diff		Diff Std. Error	Mean Diff	Sig. (2-tailed)	ddl	t	Sig.	F	
Superior	Inferior								
-,01402	-,27628	,06657	-,14515	,030	244	-2,180	,206	1,609	Hypothesis of equal variances
-,00244	-,28787	,07201	-,14515	,046	109,0	-2,016			Hypothesis of unequal variances

Source: Helis, 2014, p.130

Levene's Test shows **no significant difference in variances** (0.206 > 0.05), so the assumption of homogeneity is accepted, and we proceed with the T-test.

The **P-value (0.030)** associated with the T-statistic (T = -2.180) is statistically significant at the 0.05 level. Therefore, we reject the null hypothesis that there is no significant difference between the two groups.

Conclusion

The **Likert cumulative scale** is widely used in social and human research due to its simplicity and ability to standardize responses. However, many researchers misuse it by ignoring its conditions, which can reduce result accuracy and lead to analytical errors. Therefore, it is essential to use this scale with a thorough understanding of its strengths and limitations.

Sources and References

1. Achoui, M. (1992). Foundations of industrial psychology. Algeria: National Book Company.
2. Arab Planning Institute. (2005). Development bridge journal (Issue 24). Kuwait: Arab Planning Institute.
3. Helis, D. (2014). Job satisfaction and its relationship with job performance (Unpublished master's thesis). University of Abou Bekr Belkaid, Tlemcen.
4. Manu, C., & Fanny, P. (2009). Data analysis with SPSS (1st ed.). France: Pearson Education.
5. Howell, D. C. (2008). Statistical methods in human sciences (2nd ed.). Belgium: De Boeck.
6. Howell, D. C. (2009). Statistical methods in psychology (7th ed.). Canada: Cengage.
7. Fox, W. (1999). Social statistics (L. Imbeau, Trans. & Adapt.). Laval University Press.
8. Hodge, D. R., & Gillespie, D. (2003). Phrase completions: An alternative to Likert scales. *Social Work Research, 27*(1), 45–55. https://www.jstor.org/stable/42659450
9. Kusmaryono, I., Wijayanti, D., & Maharani, H. R. (2022). Number of response options, reliability, validity, and potential bias in the use of the likert scale education and social science research: a literature review. International Journal of Educational Methodology, 8(4), 625-637. https://doi.org/10.12973/ijem.8.4.625
10. Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from Likert-type scales. Journal of Graduate Medical Education, 5(4), 541–542. https://doi.org/10.4300/JGME-5-4-18
11. Ankur Joshi, Saket Kale, Satish Chandel & D. K. Pal. (2015). Likert Scale: Explored and Explained. Current Journal of Applied Science and Technology, 7(4), 396–403. https://doi.org/10.9734/BJAST/2015/14975