

A descriptive Analysis and Interpretation of Data from Likert Scales in Educational and Psychological Research

Hussain Alkharusi*

Abstract

Likert's scale is a measurement method to assess individual's opinions about specific issues. It might be either bipolar reflecting two opposite states of the responses or unipolar reflecting a regular graduation in the degree or status of the issue being measured. A lot of descriptive studies in educational and psychological research collect data measured on Likert scales. However, there are arguments among researchers about which methods are appropriate for a descriptive analysis and interpretation of data from Likert scales. The current paper aimed at discussing the nature of Likert scales and illustrating ways of descriptively analyzing and interpreting data collected using Likert scales. The paper started defining Likert's scale, presenting types of bias in Likert scales, and differentiating between Likert scale and Likert-type item. Then, three ways were suggested to interpret Likert scale data. These were using a scale of class intervals of the scale composite scores, one-sample t-test, and relative importance index. The selection of any of these ways should consider the conceptual framework of the issue being studied, nature of the sample, and reference of interpretation.

Keywords: Likert Scales; Likert Items; Educational Research; Psychology Research; Descriptive Research; Data Analysis

Likert scales are frequently used in educational and psychological research to measure attitudes, perceptions, beliefs and other issues. This is because most of the variables that are of interest in educational research are not directly observed and as such they are assessed by self-report measures using Likert rating scales. However, there are arguments among researchers about which methods are appropriate for a descriptive analysis and interpretation of data from Likert scales. The purpose of this paper was to discuss the nature of Likert scales and to suggest appropriate ways that could help educational and psychological researchers analyze and interpret data in a descriptive way.

Defining Likert's Scale

Likert's scale is a measurement method developed by Rensis Likert in 1932 to assess individuals' attitudes toward any object (Bertram, 2006). It refers to the extent to which individuals agree or disagree about the issue. Originally, the scale consisted of five

participants (Johns, 2010). Also, odd scales compared to even scales include a midpoint representing the neutral or moderate state of the issue. Further, the scales might be either bipolar reflecting two opposite states of the responses or unipolar reflecting a regular graduation in the degree or status of the issue being measured.

Based on the original work of Likert (1932), Likert scales involve the presentation of a set of items related to a certain issue. Each item assesses a unique aspect of the overall issue. An individual is asked to rate his or her feelings, perceptions, opinions, or attitudes in terms of agreement or satisfaction or frequency using a response set consisting of equally spaced numbers accompanied by approximately equally spaced anchors as shown in Figure 1. Each response category was assigned a numerical value so that the higher the value the higher the level of agreement, or satisfaction, or frequency depending on the type of items.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1	2	3	4	5
Very dissatisfied	Somewhat dissatisfied	Neither satisfied nor dissatisfied	Somewhat satisfied	Very satisfied
1	2	3	4	5
Never	Occasionally	Sometimes	Often	Always
1	2	3	4	5

Figure 1 Examples of Likert Rating Scales

alternatives or points. Other versions of the scale consisting of three, four, six, seven, nine, or more have also been used in the research. The difference between the various versions is that using less than five points might facilitate the responses for young participants whereas using more than six points might make the response process difficult for the

Likert suggested that the distances between the numbers in the response set as well as the distances between the anchors are equal. Thus, according to Stevens's (1946) theory of measurement scales, this implies an interval level of measurement. Likert proposed that the underlying phenomenon of interest is measured by aggregating responses of the

*Department of Psychology, College of Education, P.O.Box: 32 Al-Khod, P.C.: 123, Sultanate of Oman, Oman, E-mail: hussein5@squ.edu.om

individual across all the items by either taking the sum or the average of the responses. Thus, the composite scores derived from the Likert scales represent the interval level of measurement. The derivation of composite scores through summing up or averaging the values of responses to the items is based on the assumption that each item contributes equally to the composite score. When the items are negatively stated, the numerical values for the response categories are reversed in order to compute the composite score for the attribute. However, there is no need to reverse the scoring of the items when conducting the analysis at the item level.

Bias in Likert Scales

Likert scales are affected by three types of bias. These are central tendency bias, acquiescence bias, and social desirability bias. The central tendency bias occurs as a result of participants' avoidance of selecting extreme response options. The acquiescence bias occurs when participants agree with the items as presented. The social desirability bias occurs when participants strive to represent themselves in a more favorable way. In an effort to minimize these biases, it is recommended to use equal numbers of positive and negative response categories, include both positively and negatively stated items, and distinguish "undecided or no opinion" from "neutral or neither agree nor disagree" by placing the "no opinion" response option at the end of the response scale options (Dillman, 2000).

Likert Scale versus Likert-Type Item

It is worth to note the distinction between the individual Likert-type item and the overall Likert scale. The Likert scale refers to the set of at least three items aggregated (summed or averaged) together to represent a single attribute being measured. The rationale for having at least three items in the Likert scale is to have a reliable meaningful measurement of the attribute. The computed composite score for each individual reflects the level of the attribute possessed by the individual. The individual Likert-type item has a rank and that the differences between the items are not equally spaced. This distinction clearly suggests that data from individual Likert-type items might arguably be ordinal level of measurement whereas data from Likert scales are interval level of measurement. Therefore, Likert-type items can be analyzed by using frequencies, or finding the mode and the median, or applying non-parametric data analysis approaches. Likert scales can be analyzed by calculating the means and standard deviations as well as applying parametric data analysis approaches. An exception has been made for individual Likert-type items with numerical response formats of at least five categories in that means and standard deviations as well as parametric data analysis approaches can be applied under the conditions that all possible response categories have been selected and statistical

assumptions of the parametric tests are met (Harpe, 2015).

As stated above, each individual Likert-type item assesses a single aspect of the attribute of interest and as such it does not represent a measure of the overall attribute. Thus, inferences made on the basis of the individual Likert-type item might not apply to the overall attribute. Only inferences based on the aggregate Likert scale can be applied to the overall attribute. Hence, it is not recommended to carry out an individual item-by-item analysis for scales that have been originally developed and validated to be an aggregate measure of the attribute (Harpe, 2015).

Ways of Describing and Interpreting Likert Scale Scores

When using Likert scales, researchers often want to describe and interpret the average values or the total sum of the scores aggregated across the items. The analysis can be at two levels: the individual level to describe the overall response of the individual participant and the group level to describe the overall response of the group. The interpretation of the scores can be criterion-referenced or normative-referenced. The criterion-referenced interpretation involves determining cut-off scores based on previous literature, or experts' judgments, or a meaningful scale developed by the researcher. The normative-referenced interpretation involves determining cut-off scores based on the data themselves. There are a number of ways to describe and interpret the Likert scale scores.

The first way involves classifying the data into class intervals of means or sums of scores. This is based on the assumption that the variable under study is continuous in the population. The class intervals should have equal width. Each class interval will have a descriptive label corresponding to the issue being investigated such as level of agreement or disagreement, type and intensity of attitude, degree of importance...etc. The number of class intervals should be equal to the number of the original scale points.

To illustrate this process, let's assume that a researcher seeks to answer a research question about the level of optimism perceived by a sample of participants. The variable of interest here is the level of perceived optimism measured by 6 items using a 5-point Likert scale ranging from (1 = strongly disagree) to (5 = strongly agree). The participants indicate the extent to which they agree or disagree with each item so that higher numerical values of the scale response category represent stronger degree of agreement. After reversing the scoring of the negatively stated items, a composite score for each participant can be calculated in terms of either the sum or the average of the scores across all items. This composite score reflects the level of optimism perceived by the participant. Our aim is to describe the level of perceived optimism on a continuum scale from very low to very high. Thus, we need to develop

a scale of class intervals for interpreting the composite scores as follows:

1. Find the range of the possible composite scores by subtracting the minimum composite score from the maximum composite score. In this case, if the composite scores represent the total sum, then the range will be $(30 - 6 = 24)$ because there were 6 items each scored from 1 to 5 on the 5-point Likert scale. Therefore, the maximum total sum score will be $(5 \times 6 = 30)$ and the minimum total sum score will be $(1 \times 6 = 6)$. If the composite scores represent the means, then the range will be $(5 - 1 = 4)$ because the maximum total sum score $(5 \times 6 = 30)$ and the minimum total sum score $(1 \times 6 = 6)$ each are divided by the total number of items.
2. Find the width of the class intervals by dividing the range of the possible composite scores by the number of scale categories. This will be $(24 \div 5 = 4.80)$ for the summed scores and $(4 \div 5 = 0.80)$ for the average scores.
3. Construct the class intervals by adding the width to the lower limit of the first class interval, which is the minimum composite score, and so on for the subsequent intervals as shown in Table 1 which displays the 5-point scale of class intervals for interpreting the composite scores using the sum and Table 2 which displays the 5-point scale of class intervals for interpreting the composite scores using the average.

As shown in Table 1 and Table 2, each class interval has a uniform difference confirming with equally spaced distances between the numbers in the response

intervals. Further, all intervals are exhaustive meaning that all possible data values within the range of composite scores are included. Thus, the researcher can use either Table 1 or Table 2 to describe the average of composite scores reflecting the level of perceived optimism.

Sometimes, researchers might want to describe the variable of interest using fewer number of categories than the original measurement scale. For example, suppose the researcher wants to describe the level of optimism, which was measured on a 5-point Likert scale, using a triple scale: low, moderate, or high. In this case, in order to develop a triple scale of class intervals for interpreting the composite scores, we follow the same aforementioned steps except that the width of the class intervals will be found by dividing the range of the possible composite scores by the requested number of scale categories which is 3. This will be $(24 \div 3 = 8)$ for the summed scores and $(4 \div 3 = 1.33)$ for the average scores. After that, construct the class intervals by adding the width to the lower limit of the first class interval, which is the minimum composite score, and so on for the subsequent intervals as shown in Table 3 which displays the 3-point scale of class intervals for interpreting the composite scores using the sum and Table 4 which displays the 3-point scale of class intervals for interpreting the composite scores using the average. Although using fewer scale categories might facilitate the interpretation, it conceals some information at the extreme ends of the scale.

The second way to answer the descriptive research question concerning the level of optimism is to use

Table 1A 5-Point Scale of Class Intervals for Interpreting the Composite Scores Using the Sum

Interval	Interval	Midpoint	Interpretation
6-(6 + 4.80)	6-10.80	8.4	Very low level of optimism
10.81-(10.81 + 4.80)	10.81-15.61	13.21	Low level of optimism
15.62-(15.62 + 4.80)	15.62-20.42	18.02	Moderate level of optimism
20.43-(20.43 + 4.80)	20.43-25.23	22.83	High level of optimism
25.24-(25.24 + 4.80)	25.24-30.04	27.64	Very high level of optimism

Table 2A 5-Point Scale of Class Intervals for Interpreting the Composite Scores Using the Average

Interval	Interval	Midpoint	Interpretation
1-(1 + 0.80)	1-1.80	1.4	Very low level of optimism
1.81-(1.81 + 0.80)	1.81-2.61	2.21	Low level of optimism
2.62-(2.62 + 0.80)	2.62-3.42	3.02	Moderate level of optimism
3.43-(3.43 + 0.80)	3.43-4.23	3.83	High level of optimism
4.24-(4.24 + 0.80)	4.24-5.04	4.64	Very high level of optimism

set of the original 5-point Likert scale. This uniform difference is based on the assumption that all scores within the interval can be represented adequately by the interval midpoint. The difference between the midpoints of any two consecutive intervals as well as the difference between the upper limits of two consecutive intervals, or the difference between the lower limits of two consecutive intervals are all equal. Also, all intervals are mutually exclusive meaning that no data value can fall into two different

one sample t-test. In this way, the test variable's mean, namely "average composite score of optimism", is compared against a test value, which is a known or hypothesized value of the mean in the population. This test value may come from the literature, average estimations of experts in the field, or the midpoint of the composite scores' scale. If the test variable mean is significantly higher than the test value, then the level of perceived optimism is high. If the test variable mean is significantly lower than the

test value, then the level of perceived optimism is low. If there is no statistically significant difference between the two values, then the level of perceived optimism is moderate.

choice of any of these ways should take into account the conceptual framework underlying the topic being investigated, the characteristics of the sample, and the frame of reference for interpretation.

Table 3A triple Scale of Class Intervals for Interpreting the Composite Scores Using the Sum

Interval	Interval	Interpretation
6-(6 + 8)	6-14	Low level of optimism
14.01-(14.01 + 8)	14.01-22.01	Moderate level of optimism
22.02-(22.02 + 8)	22.02-30.02	High level of optimism

Table 4A triple Scale of Class Intervals for Interpreting the Composite Scores Using the Average

Interval	Interval	Interpretation
1-(1 + 1.33)	1-2.33	Low level of optimism
2.34-(2.34 + 1.33)	2.34-3.67	Moderate level of optimism
3.68-(3.68 + 1.33)	3.68-5.01	High level of optimism

Table 5An illustration of Calculating Relative Importance Index Values of a Scale Consisting of Three Dimensions

Dimension	<i>M</i>	<i>RII</i>
1. First dimension	3.67	$3.67 \div 5 = 0.73$
2. Second dimension	3.86	$3.86 \div 5 = 0.77$
3. Third dimension	3.47	$3.47 \div 5 = 0.69$

The third way to descriptively analyze Likert scale data is to use the relative importance index (RII). The RII is a descriptive method for determining the importance of each subscale in relation to other subscales within the scale. It is used for a descriptive analysis at the scale level. Its calculation involves dividing the average score of each subscale by (*H*) where *H* represents the highest numerical value of the response categories of the scale. It takes on values between ($\frac{1}{H}$) and 1. For example, consider a scale consisting of 30 items measured on 5-point Likert scale ranging from (*I = very low*) to (*5 = very high*). The items were divided into three dimensions so that each dimension consisted of 10 item. Table 5 presents the average score of each dimension along with the RII calculated by dividing the average score of each dimension by the highest numerical value of the response categories of the scale, which is 5. According to Table 5, the most important dimension contributing to issue measured by the scale is related to the second dimension followed by the first dimension.

Conclusion

In conclusion, data collected using self-report Likert scales are common in educational and psychological research. Researchers should differential between Likert scales and Likert-type items when considering data analysis and interpretation. Likert-type items which have Likert response options are descriptively analyzed using frequencies, modes, and medians. Likert scales which are composed of a number of Likert-type items can be interpreted or labelled descriptively by developing a scale of class intervals of the scale composite scores, using one sample t-test, and calculating the relative importance index. The

References

Bertram, D. (2006). Likert Scales: CPSC 681—Topic Report. *Poincare*, 1-11. Retrieved from <http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf>.
 Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). John Wiley and Sons.
 Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, 7(6), 836–850. <https://doi.org/10.1016/j.cptl.2015.08.001>
 Johns, R. (2010). *Likert items and scales*. Retrieved from https://dam.ukdataservice.ac.uk/media/262829/discover_likertfactsheet.pdf.
 Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 55.
 Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677-680. <https://doi.org/10.1126/science.103.2684.677>