



International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

<http://gssrr.org/index.php?journal=JournalOfBasicAndApplied>



Some Biases in Likert Scaling Usage and its Correction

Jonald L. Pimentel*

University of Southern Mindanao, Kabacan, Cotabato 9407 Philippines

Email: jonaldpimentel24@gmail.com

Abstract

In this paper, common biases or errors in the construction of intervals under the Likert scaling methodology in both odd and even scales are shown. Examples of researches that uses this scales especially in survey that requires perception response from the respondents will be presented, discussed and analyzed for possible biases. Suggestions for the corrections of these biases are presented in order to minimize the bias leading for a better labeling and interpretations of the results.

Keywords: Bias; difference; even; Likert scale; odd; scoring.

1. Introduction

Likert scaling is a psychometric concept commonly used in survey research that uses guide questionnaires that need responses in scales from the subject or respondents of interest. It is a widely used scale in many field of discipline most particularly in the social science research like in education and psychology. It is a scaling method in which a statement can be responded positively or negatively. This scaling was invented by a psychologist named Rensis Likert in 1932, whose purpose is to look for an effective and efficient means of describing attitudes of humans and its influences that affected them. Likert developed that scale that carried his name until this present time. This scale is also considered a rating scale. As an example, if a respondent wanted to respond a Likert questionnaire item, this respondent specifies a level of agreement or an order/ranking to a particular statement which can be describe in a description which in turn has an equivalent numerical values. Likert scale has many uses in psychology like on the study of behavior, marketing surveys as to the preferences of a product, education, medicine, nursing, finance, engineering and human study and the likes [1, 4, 8, 15].

* Corresponding author.

Further, it is used either as a total score that is when the scale is sum up for all items and in this case the result produced an index, this is used most commonly in psychology researches or a single ordinal score that is common among survey research like in marketing and business where they need to have an exact rank of a certain product. If a statement being ask on which a person need to respond of any criteria whether subjective or objective criteria, that statement is called a Likert item. Generally, Likert scaling is a measure on the level of say agreement or a disagreement on that particular statement.

Most social science researchers preferred to use response categories that are on odd scale (example three, five, seven or nine) because they are interested the scenario in the middle response. Further, psychometricians preferred to used seven and nine response categories [3] for reasons that they can stretch the interval on which they can describe the respondent's preference or view. Most common practiced on the other hand used five response categories while others used seven response categories, however according to reference [5] in his recent empirical study, the use of five or seven response categories produced slightly mean scores higher relatively to a possible maximum score and their difference is significant. Others researchers preferred to use the response categories that are even (example 4,6, 8 or 10), which are termed force-number Likert scale in some literature for some reasons that they want to escape the in between scenario of response scale like the three (neutral/undecided) in the five Likert scale. Most common in this case is the scale that have four response categories. Other data characteristics on the use of Liker scales shows a very little difference among the scale formats in terms of variation about the mean, skewness or kurtosis.

Among researchers, some controversies and confusion sometimes arises due to some several causes on the use of Likert scaling. Hence, an elaboration is needed. First is the treatment of the scale whether it is an ordinal scale or a measurement scale (ratio or interval) [13, 16]. This sometimes need an explanation and perhaps requires an assumption so the method applied is valid and consistent. In this study we include this an assumption so to avoid debates and confusion. Second question is the existence of central tendency bias due to the subject's avoidance of extreme response choice that is most respondents tend to go to the middle to escape form the outlying responses. Third is the presence of acquiescence bias that is an agreement with the given statements as presented, well this to refrain controversy. Fourth question is known as the social desirability bias that is putting a subjects or their group or establishments in a more favorable way hence the resulting conclusion is questionable due to this kind of error. The thought of designing a balance keying (an equal number of positive and negative statements) is always a challenge in the choice of selecting a scale that can minimize those biases and problems that may appears. Reference [12] also mentioned on the criticism among reviewers on the application of statistical methods which is not appropriate when dealings with Likert scales. Some of the criticism are well founded. For instance, violations to the assumptions of the use of some of parametric methods for example the analysis of variance, regression analysis and correlations which requires the normality assumptions on the distributions of the data or per see the appropriateness of using parametric statistical methods for Likert scaling. However, reference [12] argued that many studies since the 1930s consistently show that parametric studies can handle with respect to violations of these assumptions. Thus, those claims of inappropriateness are unfounded according to him hence, parametric methods can still be utilized without the concern of answers being invalidated. In this study, we do not attempt to apply any parametric statistical method but to do some trial and error method to further improve the method introduced by reference [13] that is to

enhanced specifically the use of Likert scaling for both even and odd number categories as well as its final equivalent description and interpretations. In this study, the researcher will illustrate in practice, for example the format of a typical five- point level Likert item on agreement coded with a numerical notation and a description legend being labeled. For example, we agree assigning 1 to strongly disagree, 2 to disagree, 3 to neither agree nor disagree (undecided or neutral), 4 to agree and 5 to strongly agree.

1.1 Limitation of the Study

This study probably is applicable only to those study that uses instruments that needs responses in scales either in an ordinal or maybe a measurement (interval or ratio) scales. The study assumes that the data in an ordinal scale can be considered to be in continuous scale that is either in interval or in ratio scale so that labeling a weighted mean (which is considered continuous) to its corresponding verbal description after it was computed is valid.

2. Methodology

2.1 Scoring and analysis

If items in a survey questionnaire instrument is responded using a Likert scale, each item maybe in two ways first it is analyzed separately given the response scale of the respondents or responses is totaled to create a score for a group of items. This is usually a common practice in some psychology and educational researches (the score is considered an index). One issue being raised is whether Likert items is an ordinal data or an interval data. This is always a subject of disagreement among researcher and users of this scale. In most practice however, they are considered ordinal or interval level data depending on the assumptions being considered. In this study we consider this as an assumption. Whether individual Likert items be considered as ordinal level data, many researchers regard such items only as an ordinal data because, in a special case when using only five levels, one cannot assume that respondents perceive all pairs of adjacent levels as equidistant. On the other hand, often the wording of response levels clearly considered both directions (symmetry) of its response levels assuming there is a middle category at. In that situation such item would fall between ordinal and an interval level measurement since treating the item as just merely ordinal, we would lose information. In the case of a scale where we consider an equal spacing of its response levels, we can immediately treat this case as an interval level data which we can argue is more strongly. Now, if we consider Likert responses an ordinal data, This responses can be displayed in a graph particularly bar charts or can describe in a median or mode, used as a center but not the average, or the variability using the range across quartiles but not the standard deviation since the average and standard deviation are inappropriate measures for ordinal data as mentioned by reference [6]. Other alternative for the analysis of the data is using a non-parametric tests on which it depends on the researcher's objective are the chi-square test (a categorical statistical method usually or count data), Mann–Whitney test, Wilcoxon signed-rank test, or Kruskal–Wallis test (comparison of mean ranks). If guaranteed by the Central Limit Theorem that ordinary averages of the Likert scale data behaves or can be assumed as an approximately normally distributed in which parametric statistical analysis can be performed. In psychological researches, Answers to several Likert questions are summed up (totaled) with the assumptions that all questions

use the same Likert scale and the scale is a defensible approximation to an interval scale, for which case it may be treated as an interval data that measures an unobserved (or latent) variable. Thus, parametric statistical tests such as the analysis of variance may be applied. Further, one consideration the method can be applied is to have more than five Likert item questions that are totaled.

2.2 Level of measurement

The five response categories are often believed to represent an interval level of measurement but this can only be in the case if the intervals between the scale points correspond to empirical observations in a metric sense. We bear in mind that the consideration of ordinal scales as an interval scales has a lot of controversial issues as mentioned previously, however in matter of agreement of its principle, it will be used as a basis for obtaining estimates in an interval level consideration on a continuum by applying item response models in which the data can be obtained in the form of an ordinal scale that fit the model particularly known as the polytomous Rasch model [20,21, 22],. The researcher recommends that the data has to be thoroughly checked to fulfill the strict formal axioms of the model before it is being analyzed.

3. Results and Discussions

Following reference [13], When items are Likert scaled and is assumed to have an interval measurement, often in practice, the information to all the respondents are summarized in the form of a weighted mean that is the scale is used as a weight multiplied to the frequency divided by the total frequency as to obtain the weighted average or mean. After this procedure the resulting weighted mean is being interpreted using an interval where in return had a corresponding verbal description. Most of these Likert scales are used in giving the level of agreement, level of frequency, level of importance or the likelihood of appearance in most application. To give an illustration in its use. Reference [17] in their paper, used four point Likert scales in the assessment of students’ self-level of motivation in mathematics. Seven survey question items were constructed for their assessment. From these items, one item was selected for discussion, the item “How much have you liked mathematics this year” which was answered using the four point Likert Scale with corresponding verbal description: 1. Not much, 2. A little, 3. some, 4. A lot. Now, some researchers in these kind of surveys, in practice created an interval of means in order to give interpretations for the weighted mean. Results of most studies would treat the problem as practiced by some researchers in which we will illustrate below

Table 1: Four point Likert Scale

Likert Scale	Interval	Verbal Description
1	1.00-1.50	Not much
2	1.51-2.50	A little
3	2.51-3.50	Some
4	3.51-4.00	A lot

Table 1 shows a table of a four point Likert scale with the constructed interval and its corresponding equivalent verbal description. This kind of interval is bias in the sense that the difference in the upper interval and lower interval of the first and last intervals is much lower as compared to the two middle intervals (see table 2 below). This observation will result items with the computed weighted mean considered more in the middle. This creates an unbalanced difference as we observed.

Table 2: Four point Likert Scale

Likert Scale	Interval	Difference	Verbal Description
1	1.00-1.50	0.50	Not much
2	1.51-2.50	0.99	A little
3	2.51-3.50	0.99	Some
4	3.51-4.00	0.49	A lot

To correct this problem, we attempt to reduce or eliminate the bias using trial and error, by making the difference in each interval having a uniform difference. A new interval was constructed as shown in Table 3 having a majority uniformed difference in each interval (except for one interval). This new interval may improve the description of the weighted mean.

Table 3: Four point Likert scale

Likert Scale	Interval	Difference	Description
1	1.00-1.75	0.75	Not much
2	1.76-2.51	0.75	A little
3	2.52-3.27	0.75	Some
4	3.28-4.00	0.72	A lot

The correction of the bias in the interval makes the difference “uniform” except for one and perhaps might change the descriptive interpretation of the item. Going back to the study of reference [17], As shown in the results of the overall weighted mean of the 5 student survey items by which yielded a values of 3.47 for IWB group and 3.22 for the control group. A description of this weighted mean leads to the interpretation for the verbal description of “some” for IWB group when using the table 2 but will have different description “a lot” when using the table 3 in which there is the presence of the correction. For the control group it has the same “some” interpretation when table 2 and table 3 are both used.

For the next tables which will be presented below, we will show the improvements for some even and odd scale categories specifically for the three, four (as shown above), five six, and seven response categories. Intervals with two decimal places are recommended below given their differences and possible description (Note

description can be change for appropriate description depending on the objective of the study)

Table 4: Three point Likert scale

Likert Scale	Interval	Difference	Description
1	1.00-1.66	0.66	Bad
2	1.67-2.33	0.66	Undecided
3	2.34-3.00	0.66	Good

Table 4 shows that for a three point Likert scale, we can create an interval in which their differences are uniform which is what we wanted to achieved.

Table 5: Five point Likert scale

Likert Scale	Interval	Difference	Description
1	1.00-1.79	0.79	Never
2	1.80-2.59	0.79	Rare
3	2.60-3.39	0.79	sometimes
4	3.40-4.19	0.79	Often
5	4.20-5.00	0.80	Always

Next is the table 5 above for a five point Likert scale in which an intervals were created with majority of the differences are similar except for one which is wider and has slighter difference of 0.1 among the rest.

For table 6 below, in a six point Likert scale an intervals were created with majority differences are similar except for one interval which is wider and has a difference of 0.03 among the rest.

Table 6: Six point Likert Scale

Likert Scale	Interval	Difference	Description
1	1.00-1.82	0.82	Very bad
2	1.83-2.65	0.82	Rather bad
3	2.66-3.48	0.82	Bad
4	3.49-4.31	0.82	Slightly good
5	4.32-5.14	0.82	Good
6	5.15-6.00	0.85	Rather good

Lastly, table 7, a seven point Likert scale with created intervals produced a majority of similar difference except the upper interval which is shorter but a very small difference of 0.01 among the rest. These per observation is a good table.

Table 7: Seven point Likert scale

Likert Scale	Interval	Difference	Description
1	1.00-1.85	0.85	Very bad
2	1.86-2.71	0.85	Rather bad
3	2.72-3.57	0.85	Bad
4	3.58-4.43	0.85	Neither good nor bad
5	4.44-5.29	0.85	Good
6	5.30-6.15	0.85	Rather good
7	6.16-7.00	0.84	Very good

For all the tables created above which shows the intervals, differences and verbal descriptions, These tables can be used as a guide or a tool for describing the verbal description of a computed weighted mean for presentation which can perhaps be considered valid and has less error (bias) preventing which is usually common in practice to some intervals created by some researchers whose intervals usually are concentrated at the middle of the scale leading to more items whose weighted assigned mostly in the middle verbal description.

4. Conclusions

Using this suggested interval perhaps will appropriately improve proper labeling for the descriptive interpretation of the computed weighted mean or average for any researches that considered Likert rating scales in each items for the whole responded questionnaires according to the number of response categories of the items.

5. Recommendations

The researcher recommends that more study should be done to refine the interval in each table by any statistical methods so to obtain uniform difference in each point Likert Scale table similar to column difference in table 4. It is also worth to discover the intervals for 8,9 and 10 point Likert scale.

References

- [1]. Anonymous. The Effect of Right or Left Placement of the Positive Response on Likert-Type Scales Used by Medical Students for Rating Instruction. *Education for Health*, vo1.1, pp.122, 1998
- [2]. E.R..Babbie. *The Basics of Social Research*. Belmont, CA: Thomson Wadsworth. p. 174 , 2005

ISBN 0534630367.

- [3]. J. Dawes. "Do Data Characteristics Change According to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales". *International Journal of Market Research*, vol. **50** (1), pp.61–77, 2008
- [4]. S. Grant, T. Aitchison, E. Henderson, J. Christie, S. Zare, J. McMurray and H. Dargie.. 'A Comparison of the Reproducibility and the Sensitivity to Change of Visual Analogue Scales, Borg Scales, and Likert Scales in Normal Subjects during Submaximal Exercise' *Chest*, vol.116, pp 1208-1218, 1999
- [5]. J.Jacoby, and M.S. Matell. 'Three-Point Likert Scales are Good Enough' *Journal of Marketing Research*, vol. 8, pp. 495-501, 1971
- [6]. S. Jamieson. *Likert scales: how to abuse them* Blackwell Publishing Ltd MEDICAL EDUCATION 2004; vol.38, pp.1212–1218, 2004
- [7]. G.P. Latham. *Work Motivation: History, Theory, Research, And Practice*. Thousand Oaks, Calif.: Sage Publications. p. 15, 2006 ISBN 0761920188.
- [8]. K.Lindhorst,, L. Corby, S. Roberts, S. Zeiler. 'Rural Consumers' Attitudes towards Nutrition Labeling' *Canadian Journal of Dietetic Practice and Research*, vol.68, pp.146-150, 2007
- [9]. R. Likert, "A Technique for the Measurement of Attitudes". *Archives of Psychology* vo.140: pp. 1–55, 1932
- [10]. L. S. Myers, G. Anthony, G. Glenn. *Applied Multivariate Research: Design and Interpretation*. Sage Publications. p. 20, 2005, ISBN 1412904129.
- [11]. N. Mogeey, (March 25, 1999). "So You Want to Use a Likert Scale?". *Learning Technology Dissemination Initiative*. Heriot-Watt University. Retrieved April 30, 2009.
- [12]. G. Norman.Likert scales, level of measurement and the "laws" of statistics. *Advances in Health Sciences Education*. vol.15(5), pp. 625-632, 2010,
- [13]. J.L..Pimentel. A Note on the Usage of Likert Scaling for Research Data Analysis. *USM R&D Journal*, vol. 18, (2), 2010. ISSN 0302-7937.
- [14]. W.E. Saris, I. Gallhofer, W.Van der Veld, W. A scientific Method for Questionnaire Design: SQP. University of Amsterdam, 2003
- [15]. M.Seal, M. Patient Advocacy and Advance Care Planning in the Acute Hospital Setting. *Australian Journal of Advanced Nursing*, vol.24, pp.29-37, 2007

- [16]. G.M.Sullivan, A.R. Artino. Analyzing and Interpreting Data from Likert-Type Scales, *Journal of Graduate Medical Education* vol. 5(4), pp.541-542, 2013
- [17]. B.Torff, R.,Tirotta. Interactive whiteboards produce small gains in elementary students' self-reported motivation in Mathematics. *Computers & Education* vol. 54, pp. 379-383, 2010 Elsevier Ltd.
- [18]. W.M. Trochim, (October 20, 2006). "Likert Scaling". *Research Methods Knowledge Base*, 2nd Edition. Retrieved April 30, 2009
- [19]. J.S. Uebersax, J. S. (2006). "Likert Scales: Dispelling the Confusion". Retrieved August 17, 2009.
- [20]. E. Muraki. Fitting a Polytomous Item Response Model to Likert-Type Data. *Applied Psychological Measurement*.vol.14.no.1. pp.59-71,1990
- [21] F. Samejima. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, no. 17. 1969
- [22]. G.N. Masters. A Rasch model for partial credit scoring. *Psychometrika*, vol. 47, pp.149-174, (1982)
- [23]. D.R. Hodge, D. Gillespie, Phrase Completions: An alternative to Likert Scales, *Social Work Research*, vol.27 pp. 45-55. .2003
- [24]. J.C. Westland. The information content of financial survey response data. *Financial Innovation* .vol.1. no.1. 2015