



INQUISITIVE TEACHER

A Peer Reviewed Refereed Research Journal

ONLINE ISSN-2455-5827

Volume V, Issue II, December 2018, pp. 122-127

www.srsshodhsansthan.org



Optimal Length of a Likert Scale – An Observation

Dr. Muhammad Idrees Ul Islam

Research Scholar, Department of Management studies, University of Kashmir,
Hazratbal, Srinagar, (Jammu & Kashmir) India.

Abstract

Likert scales are among the most widely used tools in social science research. Different numbers of response categories would likely affect response style, data distribution, reliability, and construct validity. There is a lack of research in deciding on the length of a Likert scale, when it comes to qualitative approach. This study is aimed at fulfilling this gap. We have used observation from another study and found that characteristics of the sample population can play a significant role in deciding the length of a Likert scale. Our findings are tentative in nature and more conclusive research is needed in future to substantiate them.

Keywords: Likert scale, length of a Likert scale, optimal Likert scales response categories, sample population characteristics.

Introduction

This paper is part of the thesis titled, “Trade Practices in a Conflict Zone- an Empirical Study”. Not enough literature is available on the optimal size of a Likert scale. The idea for this paper came while collecting data for the study. While collecting data and talking to some of the respondents the researchers realized that the responses were getting choked or the respondents were not getting enough scale options to respond as per their opinions/ preferences. After careful observation and talking to some research experts we realized that the length of the scale was good for some of the respondents, but it was not giving enough options to the other respondents. We had employed a 5-point scale for the study. Even though my respondents shared many characteristics like all of them belonged to the service sector of Kashmir, but they differed at the industry level viz., Hospitality, Healthcare and education. This made them dissimilar to each other as well.

In this paper, we are focusing on the optimal length of a Likert scale. By using a qualitative approach like observations from our research. What should be the optimal length of a Likert scale and how one should decide the length of the scale? What should be the key considerations while deciding on the optimal length of the scale? If your sample population is homogeneous than five-point or even a lower level Likert scale will do the job. But, if your sample population is heterogeneous than a bigger Likert scale preferably seven-point is needed. The homogeneity or heterogeneity of your sample population can be high as well as low. If we

consider the demographic information of the sample population than every sample is always heterogeneous, but this would be considered low level of deference among the respondents. On the other hand, whenever a sample is taken from respondents belonging to different groups like in our case from different industries instead of just one industry we say the heterogeneity if high, warranting a use of bigger Likert scale.

This is a qualitative study of nature as it makes use of the observations to draw inferences from another study. Most of the studies on the length of a Likert scale are quantitative and make use of statistical measure to decide on the optimal length. An observational study is defined as a study that is based on observed information and draws implications about the outcome of an interference using matching of subjects, regression or propensity score analysis, where the assignment of subjects into a cluster is outside the purview of the researcher (Rui, 2009). It is also defined as, “a type of non-randomized study in which the researchers do not interfere, instead they simply observe the progress of trials” (NCBI). It is a kind of study in which respondents are observed while collecting data or their responses without affecting the outcome of the study in any way.

Review of literature

Since psychologist Rensis Likert in, (1932) invented the Likert scale, people have tried to optimize the number of scale points. Scales from somewhere between just 2 points to a massive 101 points have been proposed at one time or another. It is a measuring tool or a scale used to determine the opinions, behaviours, and perceptions of individuals’ or consumers. It is an interview or set of questions used to know the agreements of respondents on a variety of items, products, and services. Likert scales are typically adopted in a market research. There are usually five possible choices used during the study which may include: strongly agree, agree, neutral, disagree and strongly agree (Bissonete, 2007). Over the years much literature has been written on the optimal length of a Likert scale keeping in view various considerations and various pros and cons. This review will try to shed some light on some of that literature. There are various benefits attached to the length of the scale both qualitatively and quantitatively.

Many articles published in the various International Journal describe the results obtained in studies that have used Likert-type scales. Such scales have advantages and disadvantages. The most significant consideration is to use a scale with at least five points (Allen & Seaman, 2007). We can always add “very” on the two sides of the five-category scale, increasing it to create a seven-category scale. Using a Likert scale with seven-points can increase the reliability test values of the data (Allen & Seaman, 2007). Researchers have also found that respondent preferences were highest on the 10-point scale, followed closely by the nine-point and seven-point scales (Preston & Colman, 2000). While conducting an online survey among Hispanics in Puerto Rico (Rivera). The researcher discovered that a substantial difference exists in the reliability of responses achieved through the Likert scale formats: seven, five and three points. The results recommend that the Likert scale of seven points is the most reliable and may be chosen while conducting studies online. Dawes, (2012) argued that the results obtained from a bigger Likert scale of 7 to 10 points are more comparable than a smaller scale, which may lead to

acquiring more information about the variables under study. He also grouped the 5 and 7 point scales together and argued that they are likely to give somewhat higher mean scores compared in relation to a ten-point scale, which will give the highest possible attainable mean score. Other researchers (Revilla *et al.*, 2014), argue that the studies which make use of AD scales (Agree, Disagree Scales) should focus on 5 categories rather than on 7, 10 or a higher category as this leads to acquiring data of lower quality.

Although psychometric properties are mostly similar among a different number of points, the 4-point Likert scale is not recommended for its higher skewness and lower loadings; the 11-point Likert scale from 0 to 10 is slightly preferred for its higher loadings and composite reliability (Xu & Leung, 2018). Some investigators investigated with metric using the Likert scale with no neutral point and six response categories. Avoiding a mid-point eradicates the easy way out for the respondents (Churchill & Iacobucci, 2006). While other researchers have found that adding a mid-point to rating scales improved the reliability and validity. Structural equation modeling of error structures revealed that omitting the middle alternative led respondents to randomly select one of the moderate scale points closest to where a mid-point would seem, which suggests that adding a mid-point is preferable (O'Muircheartaigh, 1999). Thus, Likert scales should have an odd number of scale points in order to comprise a mid-point as well, which, in respect of the above arguments, points to seven being the optimal length of the scale.

Researchers also show that there is no clear difference between the 2 and 6 point scales in terms of reliabilities, but a 6 point scale is preferred in respect of trace information and percentages of variance in the one-factor model and is closer to normality (Cai & Leung, 2017). Preston and Colman (2000) found that the scales with seven, eight, nine or ten points were more reliable than the scales with two, three or four points. Furthermore, they found that respondents preferred the ten-point scale, followed by the seven-point scale and the nine-point scale. Cultural differences in responses to a Likert scale have also been examined (Lee *et al.*, 2002).

The paper examines the notion of granularity in researcher-defined fixed rating scales, where granularity refers to the number of response categories or cut off points that are imposed on a scale (Smithson, 2006). The aim of this research paper is to examine the usefulness of a scale with high granularity, from the perspectives of respondents and the researcher. It is concluded that the 21-point scale was useful to respondents and by implication to researchers as well. Although researchers commonly believe using additional scale points will maximize variance, results indicate increasing scale points beyond 7 does not increase variance. Taken together, a fully labeled 7-point scale may provide the greatest benefits to researchers (Eutsler & Lang, 2015). Most of the literature reviewed above takes a quantitative approach while commenting on the optimal length of a Likert scale and forgoes qualitative approach. When it comes to deciding the optimal length of a Likert scale both approaches should be considered.

Darbyshire & McDonald, (2004) examined the direct comparison of 400 responses on 5-point and 11 –point scales to the same question, by the same people. They examine some of the important differences previously found and then illustrate the impact that has on data quality and

usability. Their conclusion, based on past research and their own analysis, is that longer, balanced and unlabelled scales offer the maximum flexibility and reliability in the majority of cases. While there are those who oppose the use of bigger scales especially 7-point and show strong support in favour of using binary scales (Dolnicara *et al.*, 2011).

Observation

As a type of closed-ended question (Dillman, *et al.* 2009), Likert type scales represent a forced choice format of questioning. Given the fixed and limited range of options provided, it is assumed that respondents would generally respond better to a wider range of meaningful choices. The question remains though, as to what can be regarded as an optimal number of response categories? It has previously been noted that the number of categories needs to be meaningful to respondents and not trivial. It is acknowledged that having the items reviewed by experts (DeVellis, 2003) and pretesting or piloting the questionnaire (Dillman, *et al.* 2009), will also facilitate the design of a questionnaire with meaningful response categories. However, this usually takes place towards the end of the design process, if it is given any purposeful and conscious attention at all. Our focus of interest here is more on the characteristics of the sample population to be considered, the design of the response categories themselves, while the questionnaire is being constructed.

Based on our observation of studying the various responses which were collected for the study. We have a major recommendation which should be considered in future studies while using Likert scales for data collection. Most of the researches in social science field use a Likert scale to measure the Opinion, Attitude, Perception, Behavior, Etc. of respondents (Likert, 1932). Our study also uses the Likert scale to measure the impact of conflict on trade practices. The question arises, is there an optimal number of responses/categories to be incorporated in every survey? Or does it differ from one study to another? We found from our observation of the respondent's behavior that if a study involves homogeneous sample population then the scale category type does not make a difference. Using a 5-point scale or a lower type is sufficient to gauge the opinion of the respondents. On the other hand, when it involves heterogeneous sample population groups as in our case respondents from various industries like Hospitality, Healthcare and Education industries participated in data gathering. Even though all are associated with the service sector of the Kashmir valley, but are very dissimilar in their business models. What we found is that using a bigger scale in this case facilitates more with inter-group comparisons than a smaller scale. As it will provide the respondents with increased choices and not to choke their responses on the scale. Preference should be given to a 7-point scale or a higher. This will stop respondents from being too neutral in their replies which seems to be the case in of smaller scales.

As we move from a sample population who are very similar in characteristics towards the sample population who have a low level of similarity in characteristics the use of more response categories is warranted. But this also is known that there is always some level of dissimilarity in our sample population in other words our sample is never completely homogeneous. Like the dissimilarity due to demographic information like age, gender, income, etc. Again this would be

considered a low-level heterogeneity as it would not affect their responses to a question much. While as, bigger dissimilarities like belonging to different industries can have a profound impact on your responses to a similar question.

Findings and future scope

Our findings from the study are exploratory in nature as we make use of our personal observations to draw inferences about the sample population. Characteristics of a sample can range from low to high in terms of similarity and high similarity warrants a use of smaller scales whereas, low level of similarity warrants a use of bigger scales. No other paper has discussed the use of sample characteristics as a precondition to deciding the length of a Likert scale. But, as we have discussed that this would have an impact on the responses and affect intergroup comparisons. In a way these are tentative findings and more conclusive research is needed in future to substantiate them. In future researchers can work on the same findings using quantitative means and statistical measure to validate our findings.

Conclusion

Likert scales containing either five, seven, nine or ten response categories have the most backing in the literature in terms of validity, reliability, discriminatory power and respondent preferences. However, despite around a century of research, different researchers may still have different opinions as to what constitutes the optimal number of scale points. There is evidence to support even smaller or bigger scales. All this is circumstantial in nature. We have tried to focus on the qualitative aspect of selecting a Likert scale and based the decision on the characteristics of the sample population.

References:

- Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality progress*, 40(7), 64.
- Bissonnette, V. L. (2007) "Statistical Tables." Victor Bissonnette's Home Page .23 Mar.2004. Dept. Of Psychology, Berry College.
- Cai, Z., & Leung, S. O. (2017). Short Forms for Attitude Scales in Social Sciences: The Case of Attitude towards Capital Punishment. *Journal of Social Service Research*, 43(5), 635-640.
- Churchill, G. A., & Iacobucci, D. (2006). *Marketing research: methodological foundations*. New York: Dryden Press.
- Darbyshire, P., & McDonald, H. (2004). Choosing response scale labels and length: Guidance for researchers and clients. *Australasian Journal of market research*, 12(2), 17-26.
- Dawes, J. G. (2012). 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. 51, No. 1, 2008.
- DeVellis, R. F. (2003) *Scale development: Theory and applications* (2nd ed.), Sage Publications, Thousand Oaks.
- Dillman, D. A., Smyth, J. D. & Christian, L. M. (2009) *Internet, mail and mixed-mode surveys: The tailored design method*, John Wiley & Sons Inc., Hoboken, N.J.

-
- Eutsler, J., & Lang, B. (2015). Rating scales in accounting research: The impact of scale points and labels. *Behavioural Research in Accounting*, 27(2), 35-51.
 - <https://www.ncbi.nlm.nih.gov/books/NBK10468/>. Accessed on 17-June-2018.
 - Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a Likert scale. *Research in nursing & health*, 25(4), 295-306.
 - Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
 - O’Muirheartaigh, C., Krosnick, J. A., & Helic, A. (1999, May). Middle alternatives, acquiescence, and the quality of questionnaire data. Paper presented at the American Association for Public Opinion Research Annual Meeting, St. Petersburg, FL.
 - Preston, C. C., & Colman, A. M. (2000). An optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1-15.
 - Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychological*, 104(1), 1-15.
 - Preston, C.C., & Colman, A.M. (2000). An optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1:15.
 - Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the number of categories in agree–disagree scales. *Sociological Methods & Research*, 43(1), 73-97.
 - Rivera, M. M. P. Is there an Optimal Number of Scale Points to Measure Attitudes and Preferences in Online Surveys among Hispanics?
 - Rui, N. (2009). Four decades of research on the effects of detracking reform: Where do we stand?—A a systematic review of the evidence. *Journal of evidence-based medicine*, 2(3), 164-183.
 - Xu, M. L., & Leung, S. O. (2018). Effects of varying numbers of Likert scale points on the factor structure of the Rosenberg Self- Esteem Scale. *Asian Journal of Social Psychology*.