International Journal of Academic Research in Management (IJARM)
Vol. 8, No. 1, 2019, Page: 1-10, ISSN: 2296-1747

www.elvedit.com

What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale

Authors

Hamed Taherdoost

hamed.taherdoost@gmail.com
Vancouver, Canada

Research Club | Hamta Group
Hamta Academy | Hamta Group
Research & Development | OBS Tech Limited
Research & Development Department | Tablokar Co

Abstract

One of the important research tool is questionnaire. Decision makers and researchers across all academic and industry sectors conduct surveys and questionnaires to uncover answers to specific, significant questions. In fact, questionnaires and surveys can be an effective tools for data collection required for research and evaluation. In order to develop a survey/questionnaire, first the researcher should decide how to collect the required data. In this regard, scaling is the branch of measurement that involves the construction of an instrument. One of the most widely used scaling method is attitude scales to measure instruments and Likert scale is applied as one of the most fundamental and frequently used psychometric tools in sociology, psychology, information system, politics, economy and many more research. However, research methodology research have not particularly suggested the best rating scale to be chosen for a research. This study is going to provide an overview of the Likert scale and comparing rating scales of different lengths. Results will make researchers able to make decision on what number of Likert scale points use for their survey and questionnaire. Taken as a whole this study suggests using of seven-point rating scale and if there is a need to have respondent to be directed on one side, then six-point scale might be the most suitable.

Key Words

Response Scale, Rating Scale, Attitude Scale, Liker Scale, Scaling Method, Response Rate, Questionnaire, Survey Design.

I. INTRODUCTION

Decision makers and researchers across all academic and industry sectors conduct surveys and questionnaires to uncover answers to specific, significant question (Taherdoost, 2016a). In fact, questionnaires and surveys are an effective tools for data collection. Once the variables of interest have been identified and defined conceptually, a specific type of scale must be selected. Scaling methods are divided into two main categories, open questions and closed question (Taherdoost, 2017b). Scaling is the process of generating the continuum, a continuous sequence of values, upon which the measured objects are placed. There are a number of factors that should be considered to choose an appropriate scaling method in a questionnaire.

An open question is one in which the respondent does not have to indicate a specific response (Taherdoost, 2017a). Open questions have a tendency to generate lengthy answers. Often, respondents see open questions as an opportunity to respond to a question in detail. As oppose to that, a closed question is one in which a respondent has to choose from a limited number of potential answers (Taherdoost, 2016b). Usually this is a straightforward yes or no. Other closed questions may require the respondent to choose from multiple response options such as multiple choice questions, Likert scale and Semantic differential scale. As articulated by Taherdoost (2017b), scale methods could be classified as a rating scales and attitude scales. Figure 1 shows some of the commonly scaling methods.

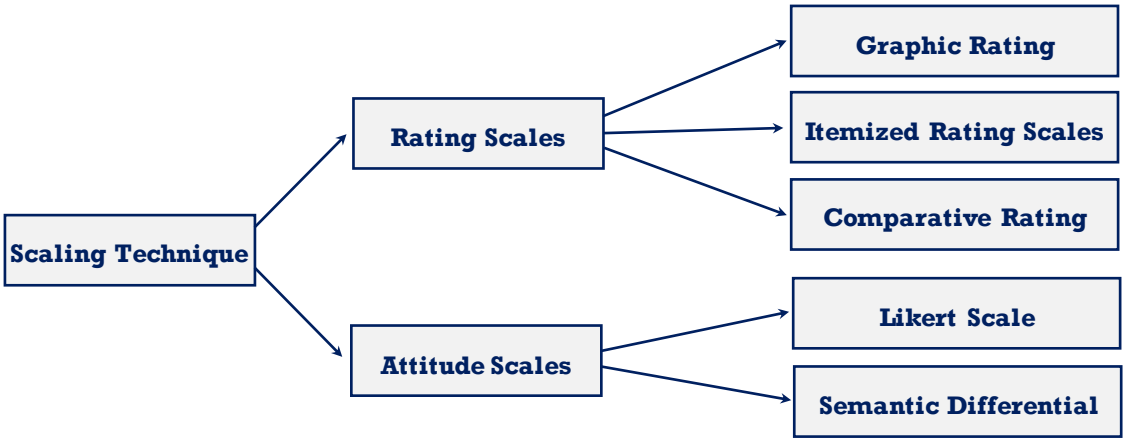

FIGURE 1: SCALING METHODS

Below is the brief description of each scaling techniques:

Rating Scales; Raters evaluate a person, object or other phenomenon at a point along a continuum or in a category. A numerical value is then assigned to this point or category. Rating scales are among the most widely used measuring instruments.

Graphic Rating Scales; Raters mark or indicate in another fashion, how they feel on a graphic scale of some sort.

Itemized Rating Scales; Raters select one of the limited numbers of categories that are ordered in some fashion. The number of categories is usually between 2 and 11.

Comparative Rating Scales; Raters judge a person, object or other phenomenon against some standard or some other person, object or other phenomenon.

Attitude Scales; Anyone of the variety of scales that measure an individual's predisposition toward any person, object or other phenomenon. These scales differ from rating scales in that they are generally more complex and multi-item scales.

Likert Scale; Respondent indicates degree of agreement and disagreement with a variety of statements about some attitude, object, person or event.

Semantic Differential; Respondent indicates how strongly he/she holds an attitude. These scales include a progression from one extreme to another (Taherdoost, 2017b).

In order to develop a survey/questionnaire, first the researcher should decide how to collect the required data (Taherdoost, 2018). In this regard, scaling is the branch of measurement that involves the construction of an instrument. There are some questions that may raise up in this step and researcher needs to release before developing the survey/questionnaire like; which scaling method should I choose for the survey/questionnaire? Does the number of response options matter? How many scales and response categories should be used? Is there an optimal number of alternatives for Likert scale items? What is the optimal number of response alternatives for a scale? What number of scale points will improve the reliability of scales? What number of scale points will improve the validity of survey? What number of scale points will increase the response rate? What number of scale points is preferred by respondents? Is there any impact of item readability if midpoint response is used? Which Likert Scale is better to use; 5-point or 7-point? Which is better; have an even or odd number of response options? Is there any advantage to use visual analog response scales than Likert scales? When should the midpoint response be endorsed in a survey?

This article is going to provide information to answer these questions by comparing rating scales of different lengths. Results will make researchers to be able to make decision on what number of rating scale points use for their survey and questionnaire. Although the review includes both scaling techniques; rating scale and attitude scale, particularly the overview is prepared to make the proper selection of Likert Scale as a technique for the measurement of attitudes. Thus rating scale, attitude scale and Likert scale may use interchangeably in this study.

II. Likert Scale

Attitude and rating scales are among the most widely used measuring instruments in like sociology, psychology, information system, politics, economy and other fields as well. However research methodology studies have not provided specific suggestion on the proper selection of rating scale for research studies (Jon A. Krosnick & Fabrigar, 1997). One of the most

fundamental and popular scaling method used in social science research is Likert scale.

Same to other scaling methods, there is debates on the number of pointes on Likert scale as well. Likert scale has been developed in 1932 as part of doctoral dissertation of Rensis Likert (Likert, 1932). This scale as a psychometric tool, includes a set of statements of research study's hypothesis. Participants in the survey are asked to state their level of agreement with those given statements from strongly agree to strongly disagree. Although the original Likert scale included five symmetrical and balanced points, during the years it has been used with different measurement range in terms of number of response options from two-points to eleven-points. Simms, Zelazny, Williams, and Bernstein (2019) summarized the Likert response labels used as shown in Table 1.

TABLE 1: LIKERT RESPONSE LABELS

Options	1	2	3	4	5	6	7	8	9	10	11
2-points	Disagree	Agree									
3-points	Disagree	Neither Agree nor Disagree	Agree								
4-points	Strongly Disagree	Disagree	Agree	Strongly Agree							
5-points	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree						
6-points	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree					
7-points	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree				
8-points	Very Strongly Disagree	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree	Very Strongly Agree			
9-points	Very Strongly Disagree	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree	Very Strongly Agree		
10-points	Very Strongly Disagree	Strongly Disagree	Disagree	Mostly Disagree	Slightly Disagree	Slightly Agree	Mostly Agree	Agree	Strongly Agree	Very Strongly Agree	
11-points	Very Strongly Disagree	Strongly Disagree	Disagree	Mostly Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Mostly Agree	Agree	Strongly Agree	Very Strongly Agree

Liker scale is simple to construct and likely to produce a highly reliable scale. Besides, from the perspective of participants, it is easy to read and complete. On the other hand, in this scale validity may be difficult to demonstrate and there is a lack of reproducibility. Additionally, another weakness of Likert scale is that participants may avoid extreme response categories and this will cause central tendency bias. Also participants may response the statements either agree or disagree in order to please the experimenter (acquiescence bias). Social desirability bias is another weakness of the Likert scale which may happen as participants may not be honest instead try to portray themselves in a more socially favorable light.

III. Comparison Rating Scales of Different Lengths

The effects of rating scales have been investigated in terms of psychometric quality criteria and systematic measurement error (Menold & Bogner, 2016). These criteria are reliability (the precision of a measurement), validity (the extent to which statements about the concepts to be measured can be made on the basis of the measurement results), response style (extreme or middle), respondent preferences and respondents-friendliness of the survey.

A. Reliability

According to Schutz and Rucker (1975) the number of response categories does not materially influence the cognitive structure derived from the results. Thus it is suggested that it has little effect on the results obtained, however information retrieval is maximized by using six or seven points (Green & Rao, 1970).

Symonds (1924) reported that inter-rater reliability is optimized using 7-point scales. Besides, (McKelvie, 1978; Nunnally, 1967) found that reliability is maximized with 7-point options. On the other hand, some researchers claimed that reliability is independent of the number of response options (Brown, Wilding, & Coulter, 1991; Matell & Jacoby, 1971).

Preston and Colman (2000) analyzed the reliability coefficients for test-retest reliability and alpha coefficients for the internal consistency reliability. They found that the highest test-retest reliability is for 7 to 10 response scales and the lowest is for 3-point. Furthermore, they reported that Cronbach alpha coefficient is highest for 11-point and with very little difference 7-point. And like test-retest reliability, the lowest is for 3-point scales. Therefore it could be concluded that reliability is increased with increasing the number of response options although from 7-point to 11-point, reliability results are all very similar.

B. Validity

Loken, Pirie, Virnig, Hinkle, and Salmon (1987) examined the criterion validity of various response categories and found that 11-point scales are superior to 3-point and 4-point scales. Oppose to above, Matell and Jacoby (1971) reported that both reliability and validity are independent of the number of scales and so by decreasing the number of response choices, reliability and validity would not be decreased.

Chang (1994) reported higher convergent validity coefficients for the 6-point scales compare to 4-point scale, however found approximately similar criterion validity for both. Preston and Colman (2000) compared scales with varying numbers of response categories in terms of criterion validity and convergent validity. According to their report, 9-point has the highest creation validity although scores from five scales to eleven point have very similar criterion validity. Their results showed that the scales with relatively more response categories (six or more) have higher convergent validity. Altogether, by increasing the numbers of scale points, validity will increase.

C. Response Preference

Jones (1968) studied the respondents' preferences for 2-point and 7-point scales and reported that respondents expressed that the 2-point scales are easier to use though the 7-points are more accurate, interesting and ambiguous. Jones (1968) concluded that respondents clearly preferred multiple-category over dichotomous scales.

Most recently, Preston and Colman (2000) examined the respondent preferences from the perspectives of "ease of use", "quick to use" and "express feelings adequately". In this study, respondents rated their level of preference from 0 to 100. Results prove that scales of five-points, ten-points and seven- points scored highest in respect of "ease of use". On the other hand, in conjunctions with "quick to use", shorter scales received the highest preference score. three-point, two-point and four-point rating scales were the most preferred. Oppose to previous two criteria, in regards to "express feeling adequacy", rating scales with more options obtained higher rating from respondents. (Preston & Colman, 2000) concluded that respondent preferences were the 10-point scale, closely followed by the 7-point and 9-point scales.

D. Odd or Even Number of Response Options

Another issue that have gotten researchers attention to develop the rating scales is if attitude and rating scales should include an even or odd number of response options (Kulas & Stachowski, 2013; Nadler, Weston, & Voyles, 2015). J.A. Krosnick (1991) suggested using midpoint scales. He mentioned that participants who wish to satisfice will look for a way to do so and if it is not obvious for them then they will choose the optimize one. He concluded that if the midpoint is not provided, then respondent will choose the optimized one however scales with middle alternative may discourage respondents from taking side in one direction. In brief, he claimed that although scales with midpoint have lower reliability, it will facilitate to collect more useful data. Additionally, according to Colman and Norris (1997), odd numbers of response categories have generally been preferred to even numbers because they allow the middle category to be interpreted as a neutral point which will give option to a person who truly has neutral position and will prevent forcing to take a side. On the point of view, there is no recommendation regarding the choice of scale and it has no effect on psychometric measurement quality criteria. Thus researchers can arrange rating scales either in ascending or descending order (Menold & Bogner, 2016).

E. Visual Analog Response Scales

Visual analog scales (Flynn, van Schaik, & van Wersch, 2004) are continuous measurement type. According to Simms et al. (2019), there is no psychometric advantage for visual analog scales rather than traditional rating scales. However non-task-related graphical elements like colors, shading or symbols should be used with caution in scales because they may affect respondents' choice.

IV. Discussion and Conclusion

According to Preston and Colman (2000), indices of reliability, validity, and discriminating power were significantly higher for scales with more response categories, up to about 7 although internal consistency did not differ significantly between scales. On the other hand, respondent preferences were highest for the ten-point scale, closely followed by the 7-point option (Preston & Colman, 2000). Besides, Miller (1956) argued that the human mind has a span of absolute judgment that can distinguish about seven distinct categories, a span of immediate memory for about seven items, and a span of attention that can encompass about six objects at a time, which suggested that any increase in number of response categories beyond six or seven might be futile.

Although Matell and Jacoby (1971) argued that the number of response options do not affect reliability and validity but some studies showed that reliability increases from 2-point to 6-point or 7-point scales (Nunnally, 1967; Symonds, 1924). Besides, studies prove that validity is increased with six or more response scales (Chang, 1994; Hancock & Klockars, 1991; Preston & Colman, 2000). Furthermore, according to Preston and Colman (2000), five-point, seven-point and 10-point scales are relatively easy to use. Although shorter rating scales are rated as relatively quick to use, scales with 10 and 11 alternatives were much preferred to express respondents feelings adequately. They concluded that 10-point, 9-point and 7-point scales are the most preferred rating scales (Preston & Colman, 2000). More to the point, rating scales that are too short cannot reveal much about the distinctions a person makes among a large set of objects, consistence with this notion, number of studies showed that longer scales conveyed more useful information up to 7-point to 9-point (Bendig, 1954) and information transfer appears to decrease for scales of 12-point of longer (McRae, 1970).

Colman and Norris (1997) mentioned that the majority of rating scales, Likert-scales and other attitude scales contain either five or seven response alternatives. Lewis (1993) concluded that 7-point scales correlate more strongly with observed significance level than 5-point scales. Besides, Finstad (2010) pointed out that seven-point scales are more likely to reflect respondents' true subjective evaluation of a usability questionnaire item than five-point options. Although Bouranta, Chitiris, and Paravantis (2009) suggested that 5-point rating scales are less confusing and increase response rate, Diefenbach, Weinstein, and O'Reilly (1993) reported that seven-point item scale emerged as the best overall and were reported by respondents as the most accurate and the easiest to use. On the point of view, Simms et al. (2019) mentioned that there is a small to non-existent difference between six-point and seven-point scales.

Taken as a whole this study suggests using of seven-point rating scale and if there is a need to have respondent to be directed on one side, then six-point scale is the most suitable.

ACKNOWLEDGMENT

This research was prepared under support of Research and Development Department of Tablokar Co and Research Club & Hamta Academy | Hamta Group.

REFERENCES

[1] Bendig, A. W. (1954). Transmitted information and the length of rating scales. *Journal of Experimental Psychology, 47*(5), 303-308.

[2] Bouranta, N., Chitiris, L., & Paravantis, J. (2009). The relationship between internal and external service quality. *International Journal of Contemporary Hospitality Management, 21*(3), 275-293.

[3] Brown, G., Wilding, R. E., & Coulter, R. L. (1991). Customer evaluation of retail salespeople using the SOCO scale: A replication extension and application. . *Journal of the Academy of Marketing Science, 9*, 374-351.

[4] Chang, L. (1994). A psychometric evaluation of four-point and six-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*, 205-215.

[5] Colman, A. M., & Norris, C. E. (1997). Comparing rating scales of different lengths: equivalence of scores from 5-point and 7-point scales. *Psychological Report, 80*, 355-362.

[6] Diefenbach, M. A., Weinstein, N. D., & O'Reilly, J. (1993). Scales for assessing perceptions of health hazard susceptibility. *Health Education Research, 8*, 181-192.

[7] Finstad, K. (2010). Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales. *Usability Metric for User Experience, 5*(3), 104-110.

[8] Flynn, D., van Schaik, P., & van Wersch, A. (2004). A comparison of multi-item Likert and visual analogue scales for the assessment of transactionally defined coping function. *European Journal of Psychological Assessment, 20*, 49-58.

[9] Green, P. E., & Rao, V. R. (1970). Rating scales and information recovery: How many scales and response categories to use? *Journal of Marketing, 34*(33-39).

[10] Hancock, G. R., & Klockars, A. J. (1991). The e□ect of scale manipulations on validity: targeting frequency rating scales for anticipated performance levels. *Applied Ergonomics, 22*, 147-154.

[11] Jones, R. R. (1968). Differences in response consistency and subjects' preferences for three personality inventory response formats. . *Proceedings of the 76th Annual Convention of the American Psychological Association*, 247-248.

[12] Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

[13] Krosnick, J. A., & Fabrigar, L. R. (1997). Designing Rating Scales for Effective Measurement in Surveys. In L. Lyberg, P. Biemer, M. Collins, E. D. Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey Measurement and Process Quality*: Wiley Series in Probability and Statistics.

[14] Kulas, J. T., & Stachowski, A. A. (2013). Respondent rationale for neither agreeing nor disagreeing: Person and item contributors to middle category endorsement intent on Likert personality indicators. . *Journal of Research in Personality, 47*, 254-262.

[15] Lewis, J. R. (1993). Multipoint scales: Mean and median differences and observed significance levels. *International Journal of Human-Computer Interaction, 5*(4), 383-392.

[16] Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology, 22*(140), 1–55.

[17] Loken, B., Pirie, P., Virnig, K. A., Hinkle, R. L., & Salmon, C. T. (1987). The use of 0-10 scales in telephone surveys. *Journal of the Market Research Society, 29*(3), 353-362.

[18] Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study 1: reliability and validity. . *Educational and Psychological Measurement, 31*, 657-674.

[19] McKelvie, S. J. (1978). Graphic rating scales: How many categories? *British Journal of Psychology, 69*, 185-202.

[20] McRae, A. W. (1970). Channel capacity in absolute judgment tasks: An artifact of information bias? *Psychological Bulletin, 73*, 112-121.

[21] Menold, N., & Bogner, K. (2016). Design of Rating Scales in Questionnaires. *GESIS Survey Guidelines. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences.*, doi: 10.15465/gesis-sg_en_15015.

[22] Miller, G. A. (1956). The magical number seven, flUS or minus two: some limits on our capacity for processing information. *Psychologica Review, 63*, 81-97.

[23] Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *Journal of General Psychology, 142*, 71-89.

[24] Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

[25] Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(2000), 1-15.

[26] Schutz, H. G., & Rucker, M. H. (1975). A comparison of variable con®gurations across scale lengths: an empirical study. *Educational and Psychological Measurement, 35*, 319-324.

[27] Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the Number of Response Options Matter? Psychometric Perspectives Using Personality Questionnaire Data. *Psychological Assessment*, 1-9.

[28] Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. *Journal of Experimental Psychology, 7*, 456-461.

[29] Taherdoost, H. (2016a). How to Design and Create an Effective Survey/Questionnaire; A Step by Step Guide. *International Journal of Academic Research in Management, 5*(4), 37-41.

[30] Taherdoost, H. (2016b). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research in Management, 5*(3), 28-36.

[31] Taherdoost, H. (2017a). Determining Sample Size; How to Calculate Survey Sample Size. *International Journal of Economics and Management Systems, 2*, 237-239.

[32] Taherdoost, H. (2017b). Measurement and Scaling Techniques in Research Methodology; Survey / Questionnaire Development. *International Journal of Academic Research in Management, 6*(1), 1-5.

[33] Taherdoost, H. (2018). Development of an adoption model to assess user acceptance of e-service technology: E-Service Technology Acceptance Model. *Behaviour & Information Technology, 37*(2), 173-197.

AUTHORS' BIOGRAPHY

Hamed Taherdoost holds PhD of Computer Science from UTM, Master of Information Security, and Bachelor degree in the field of Science of Power Electricity. With over 19 years of work experience in both industry and academic sectors, he has been involved in development of several projects in different industries including oil and gas, laboratory, hospital, transportation and IT development as a project manager, business manager, technology head, R&D head, and team leader. Apart from his experience in industry background, he also has numerous experiences in academic environment. His views on science and technology have been published in leading publications such as Elsevier, Springer, Emerald, IEEE, IGI Global, Inderscience, Taylor and Francis and he has authored over 120 scientific articles in authentic journals and conferences, seven book chapters and six Books in the fields of technology and research methodology. His current papers have been published in Behaviour & Information Technology, Information and Computer Security, Annual Data Science, Cogent Business & Management, Procedia Manufacturing & International Journal of Intelligent Engineering Informatics. Besides, he serves many international journals as editor & advisory board, and also has organized and chaired numerous conferences and conference sessions respectively. Currently, he is Senior Member of the IEEE and IEDRC, member of ISSA, AAST, SEI, IASED, and many other professional bodies. He is Certified Cyber Security Technologist, CHFI, CIS, CAPM, CEH, CISA & CISM. Currently, he is the Team Leader & Supervisor - Research & Development Manager of Research Club and Team Leader & Business Advisor of Hamta Academy | Hamta Group, Canada, Hamta Business Solution Sdn Bhd, Malaysia, and R&D Head at Tablokar Co | Switchgear Manufacturer, Iran.

Bendig, A. W. (1954). Transmitted information and the length of rating scales. *Journal of Experimental Psychology, 47*(5), 303-308.

Bouranta, N., Chitiris, L., & Paravantis, J. (2009). The relationship between internal and external service quality. *International Journal of Contemporary Hospitality Management, 21* (3), 275-293.

Brown, G., Wilding, R. E., & Coulter, R. L. (1991). Customer evaluation of retail salespeople using the SOCO scale: A replication extension and application. . *Journal of the Academy of Marketing Science, 9*, 374-351.

Chang, L. (1994). A psychometric evaluation of four-point and six-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*, 205-215.

Colman, A. M., & Norris, C. E. (1997). Comparing rating scales of different lengths: equivalence of scores from 5-point and 7-point scales. *Psychological Report, 80*, 355-362.

Diefenbach, M. A., Weinstein, N. D., & O'Reilly, J. (1993). Scales for assessing perceptions of health hazard susceptibility. *Health Education Research, 8*, 181-192.

Finstad, K. (2010). Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales. *Usability Metric for User Experience, 5*(3), 104-110.

Flynn, D., van Schaik, P., & van Wersch, A. (2004). A comparison of multi-item Likert and visual analogue scales for the assessment of transactionally defined coping function. *European Journal of Psychological Assessment, 20*, 49-58.

Green, P. E., & Rao, V. R. (1970). Rating scales and information recovery: How many scales and response categories to use? *Journal of Marketing, 34* (33-39).

Hancock, G. R., & Klockars, A. J. (1991). The e□ect of scale manipulations on validity: targeting frequency rating scales for anticipated performance levels. *Applied Ergonomics, 22*, 147-154.

Jones, R. R. (1968). Differences in response consistency and subjects' preferences for three personality inventory response formats. . *Proceedings of the 76th Annual Convention of the American Psychological Association*, 247-248.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing Rating Scales for Effective Measurement in Surveys. In L. Lyberg, P. Biemer, M. Collins, E. D. Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey Measurement and Process Quality*: Wiley Series in Probability and Statistics.

Kulas, J. T., & Stachowski, A. A. (2013). Respondent rationale for neither agreeing nor disagreeing: Person and item contributors to middle category endorsement intent on Likert personality indicators. . *Journal of Research in Personality, 47*, 254-262.

Lewis, J. R. (1993). Multipoint scales: Mean and median differences and observed significance levels. *International Journal of Human-Computer Interaction, 5*(4), 383-392.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology, 22*(140), 1–55.

Loken, B., Pirie, P., Virnig, K. A., Hinkle, R. L., & Salmon, C. T. (1987). The use of 0-10 scales in telephone surveys. *Journal of the Market Research Society, 29*(3), 353-362.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study 1: reliability and validity. . *Educational and Psychological Measurement, 31*, 657-674.

McKelvie, S. J. (1978). Graphic rating scales: How many categories? *British Journal of Psychology, 69*, 185-202.

McRae, A. W. (1970). Channel capacity in absolute judgment tasks: An artifact of information bias? *Psychological Bulletin, 73*, 112-121.

Menold, N., & Bogner, K. (2016). Design of Rating Scales in Questionnaires. *GESIS Survey Guidelines. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences.*, doi: 10.15465/gesis-sg_en_15015.

Miller, G. A. (1956). The magical number seven, flUS or minus two: some limits on our capacity for processing information. *Psychologica Review, 63*, 81-97.

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *Journal of General Psychology, 142*, 71-89.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(2000), 1-15.

Schutz, H. G., & Rucker, M. H. (1975). A comparison of variable con®gurations across scale lengths: an empirical study. *Educational and Psychological Measurement, 35*, 319-324.

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the Number of Response Options Matter? Psychometric Perspectives Using Personality Questionnaire Data. *Psychological Assessment*, 1-9.

Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. *Journal of Experimental Psychology, 7*, 456-461.

Taherdoost, H. (2016a). How to Design and Create an Effective Survey/Questionnaire; A Step by Step Guide. *International Journal of Academic Research in Management, 5*(4), 37-41.

Taherdoost, H. (2016b). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research in Management, 5*(3), 28-36.

Taherdoost, H. (2017a). Determining Sample Size; How to Calculate Survey Sample Size. *International Journal of Economics and Management Systems, 2*, 237-239.

Taherdoost, H. (2017b). Measurement and Scaling Techniques in Research Methodology; Survey / Questionnaire Development. *International Journal of Academic Research in Management, 6*(1), 1-5.

Taherdoost, H. (2018). Development of an adoption model to assess user acceptance of e-service technology: E-Service Technology Acceptance Model. *Behaviour & Information Technology, 37*(2), 173-197.