ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/366425620

Likert Scale in Social Sciences Research: Problems and Difficulties

Article *in* FWU Journal of Social Sciences · January 2023

DOI: 10.51709/19951272/Winter2022/7

CITATIONS
135

READS
53,487

3 authors:



Benidiktus Tanujaya
State University of Papua
49 PUBLICATIONS **932** CITATIONS
SEE PROFILE



Rully Charitas Indra Prahmana
Universitas Ahmad Dahlan
220 PUBLICATIONS **4,018** CITATIONS
SEE PROFILE



Jeinne Mumu
State University of Papua
33 PUBLICATIONS **800** CITATIONS
SEE PROFILE

All content following this page was uploaded by Benidiktus Tanujaya on 07 January 2023.

The user has requested enhancement of the downloaded file.

FWU Journal of Social Sciences, Winter 2022, Vol 16, No.4, 89-101
DOI: *http://doi.org/10.51709/19951272/Winter2022/7*

Likert Scale in Social Sciences Research: Problems and Difficulties

Benidiktus Tanujaya
Universitas Papua, Manokwari, Indonesia

Rully Charitas Indra Prahmana
Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Jeinne Mumu
Universitas Papua, Manokwari, Indonesia

The Likert scale is one of the essential rating scales used as a measurement tool in social sciences research, especially in the qualitative approach. Unfortunately, this scale has a great deal of controversy surrounding how data is obtained from Likert questionnaires and the appropriate statistical analysis of these data. A systematic review was performed to address this issue. Research publications from various recognized national and international articles served as research objects. This paper provides a comprehensive study of the two-perspective of the rating scales based on measurement experts, statisticians, education researchers, and other practitioners. The experts' opinions, analyses, suggestions, and solutions are obtained from journal articles, proceedings, theses, and books. After reading this article, the reader should be able to know that the accurate Likert scale produces data intervals for social sciences research. However, some requirements must be considered, specifically the composite score, midpoint, and the number of points. If these conditions are implemented, statistical methods, parametric and non-parametric, can be used to analyze the data depending on the research purpose.

***Keywords**:* interval data, Likert scale, measurement, parametric and non-parametric, social sciences research

One of the most perplexing aspects of social sciences research is that many of the variables studied are psychological or latent constructs that cannot be directly observed (Harwell & Gatti, 2001; Blanchard, Artino, & Visintainer, 2014). Latent constructs or latent variables may be hypothetical variables, unmeasurable, or unobservable (Bollen, 2002).

Quantification of a latent variable is a complicated procedure based on the measurement of selected indicators that reflect its specific features (Lautre & Fernandez, 2004). To measure a latent variable, a researcher needs a manifest variable (Harwell & Gatti, 2001) that is measured using a rating scale or rubrics. Self-confidence, motivation, and anxiety are examples of latent variables. These variables should be assessed by self-report, using a rating scale. The aim of a

Correspondence concerning this article should be addressed to Dr. Rully Charitas Indra Prahmana, Professor in Mathematics Education Department, Universitas Ahmad Dahlan, Yogyakarta, Indonesia. Email: rully.indra@mpmat.uad.ac.id

rating scale is to provide an objective response that includes and excludes responses (Garland, 1991; Albaum, 1997). One of the most widely used rating scales is the Likert Scale.

The Likert scale is widely recognized as an easy and reliable scaling technique (Royeen, 1985) with which it is easy to measure and understand respondent perception (Subedi, 2016). Van Laerhoven et al., (2004) stated that respondents found a Likert scale to be more manageable and faster than other scales. However, there is uncertainty surrounding how researchers should analyze data derived from Likert scales (Knapp, 1990; Subedi, 2016). Examples include whether the data are interval or ordinal (Cohen, Manion, & Morrison, 2000; Blaikie, 2003); whether a mid-point uses (Garland, 1991; Cox, 1980); and which words should be used as a mid-point (Singh, 2006); how many points are optimal to construct the scale; and whether parametric statistics are appropriate for questionnaire data (Harpe, 2015).

This paper conducts a comprehensive study about the controversial questions of the Likert Scale based on measurement experts, statisticians, education researchers, and other practitioners. The expert opinions, issues, challenges, analysis, suggestions, and solutions are obtained from a variety of papers and articles that can be accessed and then affirmed by our interpretations as a team of researchers. The discussion begins with the historical development of the use of the Likert scale. Further studies regarding several controversies in the Likert scale's use will be discussed in detail by presentation the opinions of several opposing experts.

Method

The purpose of this study is to conduct an analysis of published publications discussing the Likert Scale, focusing on five issues but not limited issues regarding: composite scores, midpoints, statistical controversy, number of points, and item development. We searched Google, Google scholar, and Microsoft Academic, using the search term "Likert scale" for articles published from 1 January 1972 to 31 December 2020. The publications cited were drawn from journals, books, thesis, and proceedings.

Google provides more than five million search results, whereas Google Scholar produces more than one million results. Likewise, Microsoft academics produce many articles related to the Likert scale. PDFs and reputable publications as criteria were used to analyze findings in order to decrease the quantity of articles to be evaluated.

There were 58 articles found for further evaluation. These articles are the result of research in various fields of study, especially in the social field. Articles that are not research results are also used, if they are studies based on research articles. This is so that the conclusions obtained are based on facts, not opinions.

We also manually searched specialized publications and citations from the articles identified in the original search. Once all documents were obtained, they were examined to ensure they could be used to accomplish the research objectives. Selected articles are those that address at least one aspect of the five research focuses. However, to support the study conclusions, publications that do not pertain to the five research components were also chosen. For instance, the articles that discusses the history of the Likert Scale, and which uses a Likert Scale as an instrument to collect data.

As a research strategy, we apply an integrated literature review to provide information by examining and discussing in-depth several Likert-scale studies. This procedure consists of four

steps: review design, execution, analysis, and writing (Synder, 2019; Torraco, 2005). Several key criteria, such as the topic, the aims and rationale of the study, the scope and specific research questions, and the beginning of literature collecting, are defined in the first step of review design. The researchers then conduct a literature review, beginning with determining when to do the study, analyzing and synthesizing the material, and finally testing a review technique. The method for assessing literature was critical analysis, which involves a careful evaluation of major concepts and their applicability to a situation, as well as a critique of current literature. In the interim, synthesis is conducted by combining new and current topics or ideas to generate a new formula for the subject under discussion. This study's synthesis is a novel approach to examining the subjects included in an integrated review; it is derived directly from the critical analysis and synthesis undertaken (Torraco, 2005).

In addition, the third phase is the analysis review. Reasoning and logic replace data analysis. In the integrative literature review, it serves as the conceptual foundation for arguments and explanations. These are the two most important components used to develop the proposed concept. A review that is written with precision and clarity comes last. The study's significance and necessity are addressed.

Moreover, the review procedure, including how the resource is identified, synthesized, processed, and reported by researchers, is discussed in detail. In integrated review research, review outcomes are not analyzed and evaluated as extensively as in empirical investigations. Nevertheless, the quality of a paper is evaluated by its breadth, depth, and contribution to a topic or subject that is truly original and of high value (Synder, 2019).

Results and Discussion

Brief History and Development of Likert Scale

The Likert scale was created by the sociologist Rensis Likert who proposed a scale for measuring attitudes (Likert, 1932). Likert scales measure attitudes by asking people to indicate how strongly they agree or disagree with a series of statements about a topic (McLeod, 2019). It consists of a series of related Likert-type statements about the specific attitudinal to be measured (Desselle, 2005; Willits, Theodori, & Luloff, 2016) and several items that are essential in the measurement of an underlying construct (Holt, 2014) as shown in Figure 1. Five points is a type of Likert scale that is most widely used in various social studies that use a Likert scale. It is also easier to use because it can map respondents' choices into five clear options, ranging from strongly disagree to agree strongly. If the number is seven or even 11, it will be difficult for respondents to make choices. In addition, 5 points is the number of points developed by Rensis Likert (Edmondson, 2005).

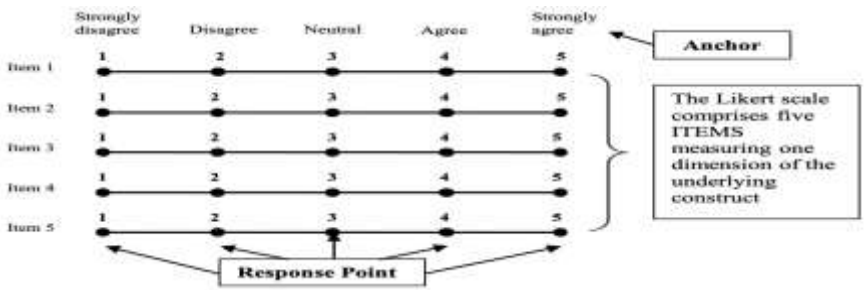

Figure 1. Structure of A Likert Scale (adapted from Holt (2014))

A Likert scale refers to a set of four or more Likert-type items combined into a single composite score used for data analysis. Typically, the researcher only wants to look at the composite score (Clason & Dormody, 1994). The items are questions or statements (Table 1).

Table 1
The Statements to Create a "Desire to Success in Mathematics" Likert Scale

No	Items	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	I study mathematics everyday					
2	I learn mathematics without being instructed					
3	I solve difficult mathematics to finish					
4	I am doing homework by myself					
5	I was very enthusiastic in mathematics class					

A Likert-type scale uses a single item that references a single Likert response. Alternatively, multiple questions may be used, but the researcher has no attempt to combine them into a single-item scale (Clason & Dormody, 1994). Table 2 provides an example of five common Likert-type items. Each item measures a distinct phenomenon and the items could not be combined to form a scale (Subedi, 2016). The five items are not part of a dimension that measures one phenomenon but consist of five different aspects.

Table 2
The Statements to Create a "Mathematics Homework" Likert-type Scale

No	Items	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	Homework is a problem for students					
2	Homework important in mathematics learning					
3	I never solve mathematics homework at home					
4	My parent helps me when I'm doing the homework					
5	Homework must be provided on every lesson					

Composite Score

Latent variables are hidden variables and thus cannot be measured directly. For example, to assess a student's motivation to learn, the teacher or researcher needs to measure several indicators, such as the student's frequency of attending classes, punctuality, and whether the student actively asks questions in class. The value of each indicator is then added up or averaged to obtain a student's level of motivation in learning. This is the composite score.

Each respondent's rating is scored as Strongly Agrees = 1, Agrees = 2, Neutral = 3, Disagrees = 4, and Strongly Disagrees = 5. An individual's score is determined by the sum of all possible points (Gay, Mills, & Airasian, 2009). The Likert scale is constructed by summing up the defining items (Michalopoulou & Symeonaki, 2017). The Likert scale is composite or 'Battery' of multiple Likert items (Johns, 2010).

An individual item is not the best measure for the entire phenomenon of interest (Harpe, 2015). Several different scales measure the phenomenon. Variables are not only measured by a collection of objects but by a collection of items that vary in size. Likert's original thinking

suggests that the intensity of the phenomena of interest be measured by the aggregate of items on the scale, not by any single entity. Separating the items generates a new scale, undermining the theory on which the original aggregated scale was developed (Harpe, 2015).

The Likert scale is a total of multiple Likert items (Brown, 2011; Boone & Boone, 2012) and can be treated as continuous data (Harpe, 2015) because human thinking and feelings are continuous variables (Yusoff & Janor, 2014). Thus, the Likert scale is also known as the summated rating scale (Royeen, 1985) or summative scale (Viljoen, 2015).

How many items are ideal for constructing a Likert scale?

Determination of the required number of items to be included in the Likert scale remains problematic (Willits, Theodori, & Luloff, 2016). Scales with too many items can cause fatigue or response distortions (Anastasi, 1976) and may also take more time to construct (Carmines & Zeller, 1979). There are no established guidelines on the number of items to be included in the final scale. Diamantopoulos et al., (2012) believe that the Likert scale should consist of at least four items. The reliability of various items increases as the number of items increases, but each item's reliability decreases as the number of items increases (Carmines & Zeller, 1979; Hinkin, 1995). Consequently, Hinkin (1998) proposed five items as the most suitable for most constructs (Hinkin, 1998).

Other types of Likert scale can be used to measure frequency, importance, quality, and likelihood (McLeod, 2019). The terms to measure frequency are: never, rarely, sometimes, often, and always (Sullivan & Artino Jr., 2013) or never, seldom, occasionally, frequently and always (Prince, 2017) (Table 3). This type is also known as a behavior rating scale.

Table 3

A Behavior Rating Scale for the Measurement of Student Participation in Learning Mathematics

No	Items	Never	Seldom	Occasionally	Frequently	Always
1	I study mathematics everyday					
2	I learn mathematics without being instructed					
3	I solve difficult mathematics to finish					
4	I am doing homework by myself					
5	I was very enthusiastic in mathematics class					

Midpoint Problems

In the use of the Likert scale, it is crucial to word a midpoint appropriately (Singh, 2006). Words used for midpoints include neutral, no opinion, undecided, and neither (Guy & Norvell, 1977; Armstrong, 1987). These words have different functions and roles and will elicit different responses.

Using the midpoint of "neither agree nor disagree" means you have a neutral opinion. Regardless of whether Undecided should be used as a midpoint, it is questionable if it is an actual midpoint between disagreement and agreement or whether it should be treated as the absence of opinion (Chyung et al., 2017). Moreover, the use of neutral and undecided is not the same. It is assumed by most that "neutral" is the neutral middle between "agree" and "disagree," but such an assumption is open to question. Even less specific is the premise of what

"undecided" means and whether respondents in the survey are aware of different meanings of the term (Guy & Norvell, 1977). Other points of view, such as I do not know or depend, should not be used as mid-point choices (Chyung et al., 2017). Therefore, the most appropriate word to use is neutral.

Controversy exists about whether the midpoint on a traditional Likert scale is worthless. Chyung's work shows that using midpoint allows neutral or indifferent respondents to express their true opinions on the topic. The respondents chose to respond or declined to respond. Sometimes, the respondent may find it challenging to respond to a question if the Likert scale uses the midpoint (Chyung et al., 2017). Respondents who choose the neutral option may indicate failure to cooperate, apathy, or laziness. An alternative possibility is that some participants have no opinion regarding the rated object (Viljoen, 2015). They do not have enough background knowledge about the topic and are ambivalent about the issue. They believe that their choice depends on other variables (Chyung et al., 2017).

Statistics Controversy

Harpe (2015) argued that the Likert scale is one of the most fundamental and often used measurement tools in the social sciences. Controversies exist in regard to appropriate data analysis. When the data are ordinal, non-parametric statistics are most appropriate for the analysis (Martilla & Carvey, 1975; Vigderhous, 1977; Knapp, 1990; Kuzon, Urbanchek, & McCabe, 1996; Jakobsson, 2004; Jamieson, 2004). If the data are measured continuously, parametric statistics may be applicable (Bishop & Herron, 2015). Some Authors state that Likert data are in the form of interval levels, so algebra can be used to calculate means and standard deviations, and the data can be analyzed using parametric statistical methods (Norman, 2010; McLeod, 2019).

Parametric analysis results are more statistically reliable than non-parametric analysis results (Bishop & Herron, 2015). Furthermore, Sullivan and Artino Jr. (2013) stated that non-parametric statistics have less power than the parametric test when tested on large samples. Parametric statistics has also a variety of methods so that researchers can use it for a variety of research purposes. However, the use of parametric statistics requires some basic assumptions. For example, for the analysis of variance - the F test, requires the assumption of normality, homogeneity, and independence.

Bishop and Herron (2015) define statistics as not substitutes for thinking about what survey data tell us and how it relates to the population. Statistical analysis is simply a method for researchers to reach a conclusion based on the data. However, it cannot draw statistical conclusions without accurate statistical analyses (Bishop & Herron, 2015).

The act of using parametric statistics for ordinal data has been described as committing the "first of the seven deadly sins of statistical analysis" (Kuzon, Urbanchek, & McCabe, 1996). Parametric statistics require data that fit into categories like ratios and intervals (Jamieson, 2004). Means and standard deviations are not always evident when descriptive terms are used in a Likert item (Sullivan & Artino Jr., 2013), for example, what is the average of the "strongly agree" and "agree" responses?

Consider the following contradictory statements, the use of parametric statistics for ordinal data – Likert data is not a deadly sin. Data Likert – although (if) classified as ordinal data, can be analyzed using parametric statistics, for example, using logistic regression to

analyze the ordinal data. Item response theory is an example of using parametric statistical methods for ordinal data. Kuzon, Urbanchek, and McCabe (1996) propose the use of statistical methods that require data with normal distribution or other continuous distribution, such as Z test, t-test, linear regression, and analysis of variance.

Are Likert scales ordinal, not interval scales? An example of interval data is how body temperature is measured using a thermometer. Suppose, Object A has a temperature of 45°C, while Object B has a temperature of 75 °C. If the two objects are heated such that the temperature of object A becomes 50 °C, while the temperature of object B becomes 80 °C, the increase in temperature is the same for each object, but the amount of energy required to achieve this differs. Analogously, on a Likert scale, the distance between each point is the same, but the perception of respondents is different depending on their background and experience. Therefore, an assessment by several respondents is needed so that the Likert data produces interval data (see discussion of composite data).

If the aim is to calculate a common set of scores for different participants, then the assigned scale will be an interval scale. However, if the researcher wants to compare different items (composite score, Likert type scale), the scale will be nominal (Tastle & Wierman, 2007). A Likert-type scale using a total score of all items is an interval scale. By contrast, items using the Likert scale are ordinal scales (Carifio & Perla, 2008). Likert treated all data as interval data (Brown, 2011; Sarafidou & Chatziioannidis, 2013; Jain, Adil, & Ananthakumar, 2013).

Six decades have propagated common misunderstandings, misconceptions, and conceptual errors regarding "Likert scales" (Carifio & Perla, 2007; 2008). With small sample sizes, different variances, and non-normal distributions, parametric statistics can be applied to Likert data without the risk of "coming to the wrong conclusion" due to the interpretation of the results (Norman, 2010). Carifio and Perla (2008) state that Likert scales are interval in nature. In general, they may be regarded ratios if they have the appropriate anchor descriptions (Holt, 2014).

Several studies have determined that parametric and non-parametric statistical tests yield comparable results when applied to Likert data, including Pearson regression and Spearman rho regression (Tastle & Wierman, 2007), the t-test and the Mann-Whitney-Wilcoxon test (De Winter & Dodou, 2010). Classical test theory and item response theory did not yield significantly different results when applied to Likert data (Liu et al., 2022). Item response theory is a measurement approach that converts ordinal data into interval data (Harwell & Gatti, 2001).

Those who argue against the use of parametric tests for the analysis of Likert data rely on theoretical assumptions to support their position. The robustness of parametric statistics enables the use of Likert data for parametric tests, even with small sample numbers and non-normal distributions (Knapp, 1990). If the Likert data contain at least four variables, the researcher should not be concerned about employing parametric statistical analysis (Baggaley & Hull, 1983).

Two types of quantitative studies can be performed on data of the Likert type. The first category includes scoring. Responses to questions are treated as numbers on a scale. They are not averaged over the items nor subjected to a factor or latent variable analysis. A respondent's score, whether weighted or unweighted, is used to measure a common property of the item set. The second type of analysis is concerned with ranking the relative significance of a group of

objects and how it may vary depending on the qualities of the individual (Dittrich et al., 2007). For a Likert scale to provide interval data, various conditions must be met, including the number of points and the use of the scale score as a composite score.

Number of Points

The most commonly used format is a 5-point scale ranging from 1 = strongly disagree to 5 = strongly agree (Jamieson, 2004; Croasmun & Ostrom, 2011). Some researchers prefer scales with 4 points (Williams, Burt, & Hilton, 2016), 6 points (Dolnicar & Grün, 2007), 7 points (Rotter, 1972; Cohen, Manion, & Morrison, 2000), 9 points (Amoo & Friedman, 2001; Carifio & Perla, 2007), and 11 points (Schwarz et al., 1991; Leung, 2011). The expression of rating for each point of the Likert scale is provided in Table 4.

Some claim that adding more points will bring the scale ever closer to a universal system (Hodge & Gillespie, 2007; Leung, 2011; Wu & Leung, 2017). A better Likert scale will result in a more representative outcome and a closer approach to the underlying distribution. The 11-point scale has good psychometric properties and easy comprehension (Leung, 2011). The number of scale steps increases, which induces respondents to use as few valid responses as possible (Matell & Jacoby, 1972). Social workers are encouraged to use 11-point Likert scales to increase generalizability, giving the subject a 0 to 10 score. If it can grow to eleven, it can be treated as a continuous measure and means that arithmetic operations can be used (Wu & Leung, 2017).

Other authors claim that the best way for the Likert scale to be reliable and valid is to use a 7-point scale (Luzano, Garcia-Cueto, & Muniz, 2008). The seven-point scoring system has good reliability, validity, sensitivity and accuracy, and good stability (Weng, 2004).

If the number of items counted exceeds the needed minimum, the gain in reliability would be so negligible that it would not be worthwhile to examine the difference or develop the instrument (Croasmun & Ostrom, 2011). However, the number of categories that may be established based on estimates of reliability is limited (Wakita, Ueshima, & Noguchi, 2012). In addition, a greater number of response possibilities may necessitate a greater mental effort from the respondent, hence diminishing response quality (Lee & Paek, 2014), response consistency (Fox & Jones, 1998), and systematic error (Lee & Paek, 2014). a number of research have demonstrated that answer quality declines above eleven options (Bendig, 1953). Simms et al. (2019) discovered no additional psychometric benefits as the number of response alternatives increased beyond six. The optimal number was therefore between four and six (Lee & Paek, 2014).

Table 4

Expressions of Rating for Each Points of the Likert Scale

No	Number of Points	Response Point	Anchors
1	4	1, 2, 3, 4	Strongly agree, agree, disagree, and strongly disagree
2	5	1, 2, 3, 4 ,5 -2, -1, 0, 1, 2	Strongly agree, agree, neutral, disagree, and strongly disagree
3	6	1, 2, 3, 4 ,5, 6	Strongly agree, agree, slightly agree, slightly disagree, disagree, strongly disagree
4	7	1, 2, 3, 4, 5, 6, 7	Completely disagree, generally disagree,

No	Number of Points	Response Point	Anchors
		-3, -2, -1, 0, 1, 2, 3	slightly disagree, neutral, slightly agree, generally agree, completely agree
5	9	1, 2, 3, 4, 5, 6, 7, 8, 9	Extremely Disagree,
		-4, -3, -2, -1, 0, 1, 2, 3, 4	……..……., Extremely Agree
6	11	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10	Extremely Disagree,
		-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5	……..……., Extremely Agree

For the question, what should the ideal number of points be for a Likert scale? A study in India shows that the Likert scale with a 5-point scale may be preferred as it is easier to implement (Choudhury & Bhattacharjee, 2014). The use of 5 points is also confirmed by Finney and DiStefano (2006), but with some conditions. If the variables have five categories or more, and the data are approximately normally distributed, so treating the data as continuous in nature and employing Maximum Likelihood estimation for statistical analysis.

Development of Item and Respondent

To acquire relevant study findings, research employing the Likert scale must also examine the quality of the items and respondents. Items should be prepared using the most effective survey formats. It is essential to use a well-developed construct definition. Utilizing terms that the target audience understands will also be advantageous. The items require opinions, avoid harshly worded statements, and use response anchors that emphasize the construct being measured, as opposed to general agreement response anchors (Rickards, Magee, & Artino Jr., 2012; Joshi et al., 2015).

Therefore, when creating a Likert scale, researchers must pay close attention to the development procedure. According to Hair Jr. et al. (2019), there are five stages of development for attitude assessment scales, including the definition and understanding of the knowledge area, the literature study or expert interviews, theoretical/face validation, semantic validation, and statistical validation.

Conclusion

Likert data produces data intervals so that arithmetic operations can be carried out, including sums, means and standard deviations. All statistical methods, parametric and non-parametric, can be used to analyze the data, depending on the purpose of the study. Examples of statistical methods that can be used to analyze the Likert data are correlation and regression method, z test, t-test, analysis of variance, structural equational modeling, and factor analysis. However, the Likert scale produces interval data when the data generated are composite scores derived from multiple items. When developing a Likert scale, it is necessary to examine the instrument's and respondents' validity and reliability.

The Likert scale is a crucial rating scale and is very popular in social research. Its scale can measure latent variables that cannot be measured directly as physical variables. In addition, it is also easy to develop and does not require too complicated requirements. The process of measuring respondents' opinions using a Likert Scale is also straightforward to do. Respondents choose one of several alternatives presented according to their choice.

Limitations of This Study

This study has several limitations, one of which is that it is not based on our own research. The study was based on the findings of a number of academics in several social science domains, including education, health, and psychology.

Recommendations for Future Research

The implementation of the Likert scale in social science research has continued to increase in the last ten years. Most use the Likert scale as one of the essential scales in analyzing the data from their research to answer the research-posed problem. However, few pay attention to several prerequisites for its use, which can result in data bias. Finally, this research's findings uses an alternative solution as a guide in using the Likert scale in social science research. Furthermore, it is necessary to carry out further studies on applying the Likert scale to various aspects analyzed in a particular field.

References

Albaum, G. (1997). The Likert scale revisited: An alternate version. *Journal of Market Research Society*, *39*(2), 331-348. https://doi.org/10.1177/147078539703900202

Amoo, T., & Friedman, H. H. (2001). Do numeric values influence subjects' responses to rating scales? *Journal of International Marketing and Marketing Research, 26*, 41-46. Available at https://ssrn.com/abstract=2423064

Anastasi, A. (1976). *Psychological Testing* (1st Ed.). New York: Macmillan.

Armstrong, R. L. (1987). The midpoint on five-point Likert-type scale. *Perceptual and Motor Skills, 64*(2), 359-362. https://doi.org/10.2466/pms.1987.64.2.359

Baggaley, A. R., & Hull, A. L. (1983). The effect of non-linear transformations on a Likert scale. *Evaluation and Health Professions, 6*(4), 483-491. https://doi.org/10.1177/016327878300600408

Bendig, A. W. (1953). The reliability of self-ratings as a function of the amount of verbal anchoring and of the number of categories on the scale. *Journal of Applied Psychology, 37*(1), 38–41. https://doi.org/10.1037/h0057911

Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. *International Journal of Exercise Science, 8*(3), 297-302. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4833473/

Blaikie, N. (2003). *Analysing Quantitative Data*. London: Sage Publications.

Blanchard, R. D., Artino Jr. A. R., & Visintainer, P. F. (2014). Applying clinical research skills to conduct education research: Important recommendations for success. *Journal of Graduate Medical Education, 6*(4), 619-622. https://doi.org/10.4300/JGME-D-14-00443.1

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*(1), 605-634. https://doi.org/10.1146/annurev.psych.53.100901.135239

Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data". *Journal of Extension, 50*(2), 1-5. Available at https://www.joe.org/joe/2012april/pdf/JOE_v50_2tt2.pdf

Brown, J. D. (2011). Likert items and scales of measurement? *SHIKEN: JALT Testing & Evaluation SIG Newsletter, 15*(1), 10-14. Available at http://hosted.jalt.org/test/bro_34.htm

Carifio, J., & Perla, R. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of the Social Sciences, 3*(3), 106-116. Available at https://thescipub.com/abstract/10.3844/jssp.2007.106.116

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education, 42*(12), 1150-1152. https://doi.org/10.1111/j.1365-2923.2008.03172.x

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Beverly Hills: Sage.

Choudhury, S., & Bhattacharjee, D. (2014). Optimal number of scale points in Likert type scales for quantifying compulsive buying behaviour. *Asian Journal of Management Research, 4*(3), 432-440. Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.431.5244&rep=rep1&type=pdf

Chyung, S. Y., Roberts, K, Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: Use of a midpoint on the Likert scale. *Performance Improvement, 56*(10), 15-23. https://doi.org/10.1002/pfi.21727

Clason, D. L., & Dormody, T. J. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education, 35*(4), 31-35. Available at https://pdfs.semanticscholar.org/e2ad/36059a7894c 89ea566811eb82ad92acd88a7.pdf

Cohen, L., Manion, L., & Morrison, K. (2000). *Research Methods in Education*. London: Routledge Farmer. Available at https://gtu.ge/Agro-Lib/RESEARCH%20METHOD%20COHEN%20ok.pdf

Cox, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research, 17*(4), 407-442. https://doi.org/10.1177/002224378001700401

Croasmun, J. T., & Ostrom, L. (2011). Using Likert-type scales in the social sciences. *Journal of Adult Education, 40*(1), 19-22. Available at https://eric.ed.gov/?id=EJ961998

De Winter, J. C. F., & Dodou, D. (2010). Five-point Likert items: T test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation, 15*(11), 36-49. Available at http://citeseerx.ist.psu.edu/viewdoc/summary? doi=10.1.1.650.3292

Desselle, S. P. (2005). Construction, implementation, and analysis of summated rating attitude scales. *American Journal of Pharmaceutical Education, 69*(5), 97.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. *Journal of the Academy of Marketing Science, 40*(3), 434-449. https://doi.org/10.1007/s11747-011-0300-3

Dittrich, R., Francis, B., Hatzinger, R., & Katzenbeisser, W. (2007). A paired comparison approach for the analysis of sets of Likert-scale responses. *Statistical Modelling, 7*(1), 3-28. https://doi.org/10.1177/1471082X0600700102

Dolnicar, S., & Grün, B. (2007). Cross-cultural differences in survey response patterns. *International Marketing Review, 24*(2), 127-143. https://doi.org/10.1108/02651330710741785

Edmondson, D. (2005). Likert scales: A history. *Proceedings of the Conference on Historical Analysis and Research in Marketing, 12*, 127-133.

Finney, S.J. & DiStefano, C. (2006). Non-normal and Categorical data in structural equation modelling. In G. r. Hancock & R. O. Mueller (Hrsg.). *Structural equation modelling: a second course* (pp. 269–314). Greenwich, Connecticut: Information Age Publishing

Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology, 45*(1), 30–45. https://doi.org/10.1037/0022-0167.45.1.30

Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin, 2*(1), 66-70. Available at https://www.rangevoting.org/MB_V2_N3_Garland.pdf

Gay, L. R., Mills, G. E., & Airasian, P. (2009). *Educational Research: Competencies for Analysis and Applications*. Columbus, OH: Merrill. Available at http://englishlangkan.com/produk/E%20Book%20Educational%20Research%20L%20R%20Gay%20Pearson%202012.pdf

Guy, R. F., & Norvell, M. (1977). The neutral point on a Likert scale. *Journal of Psychology, 95*(2), 199-204. https://doi.org/10.1080/00223980.1977.9915880

Hair Jr., J. F., LDS Gabriel, M., Silva, D. D., & Braga Junior, S. (2019). Development and validation of attitudes measurement scales: fundamental and practical aspects. *RAUSP Management Journal, 54*(4), 490-507. https://doi.org/10.1108/RAUSP-05-2019-0098

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

Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research, 71*(1), 105-131. https://doi.org/10.3102/00346543071001105

Hinkin, T. R. (1995). A Review of scale development practices in the study of organizations. *Journal of Management, 21*(5), 967-988. https://doi.org/10.1177/014920639502100509

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104-121. https://doi.org/10.1177/109442819800100106

Hodge, D. R., & Gillespie, D. F. (2007). Phrase completion scales: A better measurement approach than Likert Scales? *Journal of Social Service Research, 33*(4), 1-12. https://doi.org/10.1300/J079v33n04_01

Holt, G. D. (2014). Asking questions, analysing answers: Relative importance revisited. *Construction Innovation, 14*(1), 2-16. https://doi.org/10.1108/ci-06-2012-0035

Jain, B., Adil, G. K., & Ananthakumar, U. (2013). An instrument to measure factors of strategic manufacturing effectiveness based on Hayes and Wheelwright's model. *Journal of Manufacturing Technology Management, 24*(6), 812-829. https://doi.org/10.1108/jmtm-11-2011-0102

Jakobsson, U. (2004). Statistical presentation and analysis of ordinal data in nursing research. *Scandinavian Journal of Caring Sciences, 18*(4), 437-440. https://doi.org/10.1111/j.1471-6712.2004.00305.x

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education, 38*(12), 1212-1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Johns, R. (2010). Likert items and scales. *Survey Question Bank*: *Methods Fact Sheet 1*. Available at: https://ukdataservice.ac.uk/media/262829/discover_likertfactsheet.pdf

Joshi, A., Kale, S., Chandel, S., & Pall, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology, 7*(4), 396-403. https://doi.org/10.9734/BJAST/2015/14975

Knapp, T. R. (1990). Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research, 3*(2), 121-123. https://doi.org/10.1097/00006199-199003000-00019

Kuzon, W. M., Urbanchek, M. G., & McCabe, S. (1996). The seven deadly sins of statistical analysis. *Annals of Plastic Surgery, 37*, 265-272. https://doi.org/10.1097/00000637-199609000-00006

Lautre, I, G., & Fernandez, E. A. (2004). A methodology for measuring latent variables based on multiple factor analysis. *Computational Statistics & Data Analysis, 45*(3), 505-517. https://doi.org/10.1016/S0167-9473(03)00037-9

Lee, J., & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment, 32*(7), 663–673. https://doi.org/10.1177/0734282914522200

Leung, S. O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research, 37*(4), 412-421. https://doi.org/10.1080/01488376.2011.580697

Likert, R. (1932). The Method of Constructing an Attitude Scale, in Fishbein, M. (Ed) (1967) *Readings in Attitude Theory and Measurement*. New York: John Willey & Sons, Inc., 90 - 95.

Liu, J., Lin, H., Hu, B., Zhou, Z., Agyeiwaah, E., & Xu, Y. (2022). Advancing the understanding of the resident pro-tourism behavior scale: An integration of item response theory and classical test theory. *Journal of Business Research, 141*, 113-125. https://doi.org/10.1016/j.jbusres.2021.12.013

Luzano, L. M., Garcia-Cueto, E., & Muniz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology, 4*(2), 73-79. https://doi.org/10.1027/1614-2241.4.2.73

Martilla, J. A., & Carvey, D. W. (1975). Four subtle sins in marketing research. *Journal of Marketing, 39*(1), 8-15. https://doi.org/10.1177/002224297503900103

Matell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology, 56*(6), 506–509. https://doi.org/10.1037/h0033601

McLeod, S. A. (2019). *Likert Scale Definition, Example and Analysis*. Retrieve from www.simplypsychology.org/likert-scale.html

Michalopoulou, C., & Symeonaki, M. (2017). Improving Likert scale raw scores interpretability with K-means clustering. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 135*(1), 101-109. https://doi.org/10.1177/0759106317710863

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*(5), 625-632. https://doi.org/10.1007/s10459-010-9222-y

Prince, L. R. (2017). *Psychometric Methods: Theory into Practice.* New York: The Guilford Press.

Rickards, G., Magee, C., & Artino Jr., A. R. (2012). You can't fix by analysis what you've spoiled by design: Developing survey instruments and collecting validity evidence. *Journal of Graduate Medical Education, 4*(4), 407-410. https://doi.org/10.4300/JGME-D-12-00239.1

Rotter, G. S. (1972). Attitudinal points of agreement and disagreement. *The Journal of Social Psychology, 86*(2), 211-218. https://doi.org/10.1080/00224545.1972.9918619

Royeen, C. B. (1985). Adaptation of Likert scaling for use with children. *The Occupational Therapy Journal of Research, 5*(1), 59-69. https://doi.org/10.1177/153944928500500104

Sarafidou, J. O., & Chatziioannidis, G. (2013). Teacher participation in decision making and its impact on school and teachers. *International Journal of Educational Management, 27*(2), 170-183. https://doi.org/10.1108/09513541311297586

Schwarz, N., Knauper, B., Hippler, H. J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly, 55*(4), 570-582. https://doi.org/10.1086/269282

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, 31*(4), 557–566. https://doi.org/10.1037/pas0000648

Singh, Y. K. (2006). *Fundamental of Research Methodology and Statistics*. New Delhi: New Age International (P) Ltd. Publisher. Available at https://epdf.pub/fundamental-of-research-methodology-and-statistics.html

Subedi, B. S. (2016). Using Likert type data in social science research: Confusion, issues and challenges. *International Journal of Contemporary Applied Sciences, 3*(2), 36-49. Available at http://www.ijcar.net/assets/pdf/Vol3-No2-February2016/02.pdf

Sullivan, G., & Artino Jr., A. R. (2013). Analyzing and interpreting data from Likert-type scales". *Journal of Graduate Medical Education, 5*(4), 541-542. https://doi.org/10.4300/JGME-5-4-18

Synder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of business research, 104,* 333-339.

Tastle, W. J., & Wierman, M. J. (2007). Consensus and dissention: A measure of ordinal dispersion. *International Journal of Approximate Reasoning, 45*(3), 531-54. https://doi.org/10.1016/j.ijar.2006.06.024

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review, 4*(3), 356-367.

Van Laerhoven, H., van der Zaag-Loonen, H., & Derks, B. H. F. (2004). A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatrica, 93*(6), 830-835. https://doi.org/10.1111/j.1651-2227.2004.tb03026.x

Vigderhous, G. (1977). The level of measurement and "permissible" statistical analysis in social research. *The Pacific Sociological Review, 20*(1), 61-72. https://doi.org/10.2307/1388904

Viljoen, M. (2015). Constructing homogeneous Likert-type summative rating scales according to classical measurement theory. *Journal of Social Sciences, 43*(2), 143-151. https://doi.org/10.1080/09718923.2015.11893431

Wakita, T., Ueshima, N., & Noguchi H. (2012). Psychological distance between categories in the Likert scale: Comparing different numbers of options. *Educational and Psychological Measurement, 72*(4), 533-546. https://doi.org/10.1177/0013164411431162

Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64*(6), 956-972. https://doi.org/10.1177/0013164404268674

Williams, K. L., Burt, B. A., & Hilton, A. A. (2016). Math achievement: A role strain and adaptation approach. *Journal for Multicultural Education, 10*(3), 368-383. https://doi.org/10.1108/JME-01-2016-0005

Willits, F., Theodori, G., & Luloff, A. (2016). Another look at Likert scales. *Journal of Rural Social Sciences, 31*(3), Article 6.

Wu, H., & Leung, S. O. (2017). Can Likert scales be treated as interval scales? - A Simulation study. *Journal of Social Service Research, 43*(4), 527-532. https://doi.org/10.1080/01488376.2017.1329775

Yusoff, R., & Janor, R. M. (2014). Generation of an interval metric scale to measure attitude. *Sage Open, 4*(1), 1-16. https://doi.org/10.1177/2158244013516768

View publication stats