



THE LIKERT SCALE: EXPLORING THE UNKNOWN AND THEIR POTENTIAL TO MISLEAD THE WORLD

^{1*}Alhassan I., ²Asiamah N., ³Opuni, F. F., ⁴Alhassan A.

¹Planning and Research Unit, University for Development Studies, Tamale, Ghana

²Department of Epidemiological Methods, Africa Centre for Epidemiology, Accra, Ghana

³Department of Marketing, School of Business, Accra Technical University, Accra, Ghana

⁴Department of Statistics, Faculty of Physical Sciences, University for Development Studies, Ghana

*Author's, Email address: ibrahim.alhassan@uds.edu.gh

Abstract

The Likert scale is such an admired and pervasive data collection tool, that has gained great recognition, especially in the last two decades. However, both seasoned and early-career researchers sometimes use it thoughtlessly in their investigations, resulting in the production of misleading research findings. This paper presents a general discourse on Likert scales and is a response to the misunderstandings surrounding their use, especially in recent years. The paper is based on secondary data and advances knowledge on research instrumentation using the Likert scale. We discussed some of the problems that can arise from the inappropriate use of the scale as well as strategies for reducing its impact on research findings. Thus, we proposed some remedies for dealing with “noisy” data obtained from a poorly constructed and guilelessly applied Likert scale. Eventually, we proposed a unified Likert scale that resolves the conflicting functions of descriptive anchors and numerical labelling of the scale.

Keywords: Attitude, Continuum, Data, Likert scale, Psychometric

Introduction

It has long been known that the level of endorsement achieved by a product or service is largely informed by the mental and neural disposition that people hold toward that product, service, or event (Rob, 2010). Psychologists and behavioural scientists have been trying to find scientific ways to measure human attitudes, as they believe that a person's attitude determines how successful he or she will be in all aspects of life, including his or her own physical and mental development (Rinker, 2014; Rob, 2010). The search for a useful and needful measuring scale has always eluded many researchers and academics in the behavioural sciences (Joshi et al., 2015). While some have placed blame on the seemingly “dishonest” nature of

research participants during the evaluation or rating process, others have cited the eclectic nature (i.e., the cognitive, affective, and behavioural components) of human attitude as the primary reason there seems to be no flawless scale to measure human attitude (Rinker, 2014).

The development, usage, and popularity of psychometric scales date back to the 1920s, when Louis Leon Thurstone developed the first formal technique to measure attitudes toward religion by asking participants to agree or disagree with a series of related statements (Jovancic, 2020). Thurstone was armed with the notion that human attitude was cumulative and that he could, therefore,

calculate it as the sum of all statements that participants rated (Roufun, 2018). Besides being credited with the pioneering effort of psychometric scaling, Thurstone was also recognized as the most renowned productive scaling theorist, following his invention of “equal-appearing intervals”, “successive intervals” and “paired comparisons.”

On a different wavelength of thought, Louis Guttman, an American mathematician and sociologist, proposed a different scale, called the Guttman Scale, to measure attitude (Roufun, 2018). Guttman’s scale measures the strength of a respondent’s opinion about an event, activity, product, or service (Jovancic, 2020).

Despite the efforts made by Thurstone and Guttman, another American psychologist, Rensis Likert, like his contemporaries, also did not agree with the then-popular belief held by most psychologists that latent phenomena like attitude could not be measured. Thus, what became known as the Likert method of measuring attitude was formulated in his doctoral dissertation, and an abridged version appeared in a 1932 article in the *Archives of Psychology* (Rob, 2010). Since then, Likert’s method of attitude measurement has remained the most widely used and popular psychometric scale in questionnaires and survey research.

As mentors and research supervisors, we have observed a noticeable increase in the use of Likert scales in most studies, especially in the last decade. We have also observed that, with such an important, admired, and pervasive data collection tool, both seasoned and early-career researchers sometimes misapplied it in their investigations. As a result, flawed research findings are produced.

Indeed, we are concerned about the small body of literature that provides the appropriate know-how for Likert scale construction and how not to misuse it. Therefore, we are of the view that studies and/or documentation (such as this current effort) focusing on explaining the Likert scale are urgently needed. Admittedly, however,

discourse on the problems relating to the use of the Likert scale has long been noticed. For instance, the Likert scale literature is flooded with studies debating the ordinal or interval nature of Likert scale data.

While some studies (Adelson & McCoach, 2010; Chang, 1994; Guy & Norvell, 1977; Leung, 2011; Preston & Colman, 2000) are also discussing the required number of descriptive anchors that a Likert scale should have to ensure the validity and reliability of its data, others (Durand & Lambert, 1988; Kulas et al., 2008; Raaijmakers et al., 2000) are busy battling one another on the labelling that should be given to the midpoint of a Likert scale.

Unfortunately, there has been virtually no discussion on how improper application of the Likert scale can result in the production of noisy data and, therefore, incorrect findings, even though some researchers (Peter & John, 2003; Chyung et al., 2017) came close to noticing this when they argued that the Likert scale’s respondents do not always interpret and use the midpoint of the scale in the way that researchers or scale developers intended and that survey participants might select a midpoint even if their true opinion is not neutral.

In view of this, our main objective is to point out the mishaps relating to the inappropriate use of the Likert scale. The paper is based on secondary data and advances knowledge on research instrumentation using the Likert scale. In this paper, we present the types of data produced by psychometric scales. This is followed by an overview of the Likert scale. We also present a discussion on how the midpoint of a Likert scale disrupts or discontinues the supposed continuum nature of a typical Likert scale, especially of the bipolar type. We then proposed a remedy should such an error occur, resulting in inappropriate use of the Likert scale, and finally, we presented best practices for handling the Likert scale going forward.

Types of Data Produced by Psychometric Scales

The origins of many of the data types used in psychometric scales can be traced back to Stevens' (1946) work "On the Scale of Measurement," which has become well-established in the literature. According to Stevens (1946), scale data are classified into nominal, ordinal, interval, and ratio. Stevens further classified this scale data into intensive (nominal and ordinal) and extensive (interval and ratio). He defined the *nominal scale* as data belonging to different categories with no clear order or zero point.

Stevens (1946) believed that the *ordinal scale* has all the properties of the nominal scale and includes a sense of order in terms of classification. *Interval scales* are similar to ordinal scales in that they include concepts and categories that are classified, but the distance between them is equidistant (Stevens, 1946). A *ratio scale* is a type of interval scale with a "true zero point." Ratio scales mean that there is order and meaningful spacing (Peter, 2018).

Inferring, the nominal data type has fewer mathematical properties than the higher

hierarchy data types. The nominal gives data on categories, the ordinal on the sequences, interval discloses the magnitude between the points of the scale, and the ratio describes the order and absolute distance between any two points on the scale.

The Likert Scale

A Likert scale is a rating scale used to assess opinions, attitudes, or behaviours (Bhandari, 2022). A Likert scale is composed of three or more statements that evaluate a single construct, usually an attitude or trait, when response scores are summed (Boone & Boone, 2012). While the original Likert scale, developed by Rensis Likert, used a scale based on the participants' approval of statements, ranging from Strongly Disapprove through Disapprove, Neutral/Undecided, and Approve to Strongly Approve (Likert, 1932), the Likert scale, as it is known and used today, often represents a series of statements about which participants are asked to indicate their degree of agreement or disagreement. As mentioned earlier, for Likert, human attitudes towards any object or an issue varied along a negative-to-positive dimension and, therefore, could be measured with the scale illustrated in Figure 1.

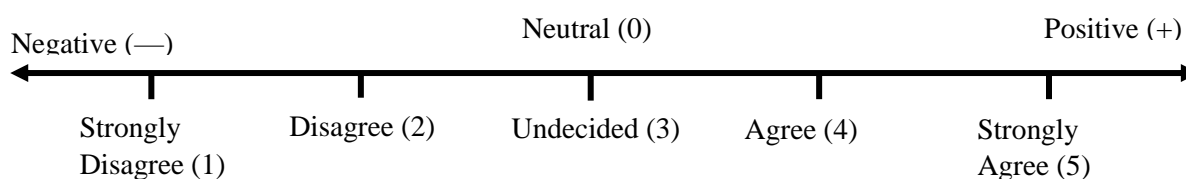


Figure 1: An Example of a Likert Scale

Source: Authors' Elaboration, 2021.

Figure 1 is an example of a Likert scale. It is represented by a horizontal line called the "visual analogue scale," on which participants are expected to indicate their degree of agreement or disagreement by circling or tick-marking a descriptive anchor that represents a participant's opinion about an item or statement.

Parts of a Likert Scale

In general, a typical Likert scale has three parts. They are the preamble, the items/indicators, and the descriptive anchors or descriptors, which come with their respective scores.

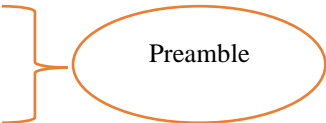
The *preamble* is an introductory statement describing what the scale measures (the construct) and what the participant should know. An example of a preamble is stated as:

“The following items or statements describe how often you experience abuse and neglect in your neighbourhood. On a scale of 1 to 3, where 1 = not at all, 2 = sometimes, and 3 = always, indicate how often you encounter people in your neighbourhood who..... ”

The *items* are the various statements or questions that are presented as indicators of the construct or the latent variable. They are the individual questions that together indicate a respondent’s overall judgement about the construct.

Descriptive Anchors are the descriptors with assigned scores that participants tick or mark to indicate their opinion about a statement. Consider Figure 2 below.

The following items or statements describe how often you experience abuse and neglect in your neighbourhood. On a scale of 1 to 3, where 1 = not at all, 2 = sometimes, and 3 = always, indicate how often you encounter people in your neighbourhood who



S/No.	Item	Not at all (1)	Sometimes (2)	Always (3)
1.	The IT system is user-friendly			
2.	The IT system facilitate my work			
3.	The IT system has robust functions			
4.	My Mentor is using the system as well			
5.	My co-workers are using the system			
6.	The electricity power system supports the system			
7.	There is regular internet connectivity for the system			
8.	The internet is strength is enough for the system			

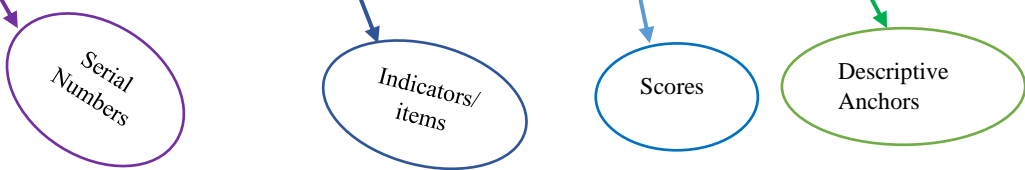


Figure 2: Parts of the Likert scale

Source: Authors’ Elaboration, 2021.

Types of Likert Scales

Likert scales can be categorized in terms of (1) the number of descriptive anchors, (2) the nature of scores, and (3) what the scale seeks to measure.

Types of Likert Scale Based on the Number of Descriptive Anchors

Table 1 illustrates the types of Likert scales on the bases of descriptive anchors

Table 1: Types of Likert Scale Based on the Number of Descriptive Anchors

Number of Points	Description	Example
3-Point Likert Scale	They have 3 descriptive anchors	1) Disagree, 2) Neutral and 3) Agree
4-Point Likert Scale	They have 4 descriptive anchors	1) Strongly Disagree, 2) Disagree, 3) Agree, and 4) Strongly Agree
5-Point Likert Scale	They have 5 descriptive anchors	1) Strongly Disagree, 2) Disagree, 3) Neutral 4) Agree, and 5) Strongly Agree
7-Point Likert Scale	They have 7 descriptive anchors	1) Strongly Disagree, 2) Disagree, 3) Somewhat disagree 4) Neutral 5) Somewhat Agree 6) Agree, and 7) Strongly Agree

Source: Authors' Elaboration, 2021.

Types of Likert Scale Based on the Nature of Scores

In terms of their nature, Likert scales are categorized into bipolar and unipolar scales.

A *unipolar* Likert scale refers to scales with numerical values extending in only one direction that seek to measure only one trait (e.g., satisfaction) in a continuum (Bhandari, 2020). Unipolar scales include descriptive anchors that can be arranged from the highest to the lowest or vice versa. There is no middle anchor that separates two opposite sets of anchors. A response category of “dissatisfied”, “moderately satisfied”, “satisfied”, and “very satisfied” is a typical example of a unipolar Likert scale since it is one-directional.

The *bipolar* Likert scale refers to a scale with numerical values ranging from positive through zero to negative values (Bhandari, 2020) as a result, they are two-directional and seek to measure two attributes (e.g., satisfaction and dissatisfaction) in a continuum. There is a middle anchor that separates the two opposite sets of anchors, with a middle (often neutral) anchor in view. Thus, the term “bipolar” is used. A typical example of a bipolar Likert scale is the response category of “Very Dissatisfied,” “Dissatisfied,” “Neutral,” “Satisfied,” and “Very Satisfied.”

Types of Likert Scale Based on What the Scale Seeks to Measure

A cursory glance at the literature will indicate dozens of variations on themes measuring attitudes such as satisfaction, likelihood, agreement, frequency, quality, and so on. Each of these scale types is worthy of elaboration. The elaboration is justified because the type of scale to be used will largely depend on what is being measured. For example, it is not appropriate to use a satisfaction scale to measure the relevance of an orientation programme. In this example, the appropriate scale is agreement.

The Satisfaction Scale: requires participants to indicate their subjective opinion about, for example, a company’s product or service based on their experience. Here, participants must consume the product or service or have a sense of the company's brand to be able to indicate their degree of satisfaction.

The Likelihood Scale: The likelihood scale, also known as the probable scale, is typically used to determine whether, for example, your customers will continue to buy a specific product or recommend your company to others.

The Agreement Scale: This scale is the most widely used scale across industries and

requires survey participants to indicate their degree of agreement or disagreement with a statement or Likert item (Joshi et al., 2015). An agreement response anchor may be cast as “Strongly disagree”, “Disagree” “Neutral”, “Agree”, “Strongly Agree”.

The Frequency Scale: this scale requires participants to indicate how frequently they perform a certain activity. We can have frequency response anchors like: “Never,” “Rarely” “Sometimes,” “Often,” “Always.”

The Quality Scale: researchers often use this type of scale when they want participants to express their perceptions about the quality of a product or service. We present below a summary of Likert-type descriptive or response anchors.

Table 2: A Summary of Likert Scale-Type Descriptive Anchors

Level of Appropriateness	Level of Importance	Knowledge of Action	My beliefs	Priority	Reflect Me
1 – Absolutely inappropriate 2 – Inappropriate 3 – Slightly inappropriate 4 – Neutral 5 – Slightly appropriate 6 – Appropriate 7 – Absolutely appropriate	1 – Not at all important 2 – Low importance 3 – Slightly important 4 – Neutral 5 – Moderately important 6 – Very important 7 – Extremely important	1 – Never true 2 – Rarely true 3 – Sometimes not true 4 – Neutral 5 – Sometimes true 6 – Usually true 7 – Always true	1 – Very untrue of what I believe 2 – Untrue of what I believe 3 – Somewhat untrue of what I believe 4 – Neutral 5 – Somewhat true of what I believe 6 – True of what I believe 7 – Very true of what I believe	1 – Not a priority 2 – Low priority 3 – Somewhat priority 4 – Neutral 5 – Moderate Priority 6 – High priority 7 – Essential priority	1 – Very untrue of me 2 – Untrue of me 3 – Somewhat untrue of me 4 – Neutral 5 – Somewhat true of me 6 – True of me 7 – Very true of m
Level of Agreement	Likelihood	Level of Satisfaction	Level of Quality	Frequency	Level of Acceptance
1 – Strongly disagree 2 – Disagree 3 – Neither agree nor disagree 4 – Agree 5 – Strongly agree	1 – Extremely unlikely 2 – unlikely 3 – Neutral 4 – likely 5 – Extremely likely	1 – Very dissatisfied 2 – dissatisfied 3 – unsure 4 – satisfied 5 – Very satisfied	1 – Poor 2 – Fair 3 – Good 4 – Very good 5 – Excellent	1 – Never 2 – Rarely 3 – Sometimes 4 – Often 5 – Always	1 – Totally unacceptable 2 – Unacceptable 3 – Neutral 4 – Acceptable 5 – Perfectly Acceptable
Level of Problem	Barriers	Good / Bad	Level of Responsibility	Comparison of Two Products	Level of Influence
1 – Not at all a problem 2 – Minor problem 3 – Moderate problem 4 – Serious problem	1 – Not a barrier 2 – Somewhat of a barrier 3 – Moderate barrier 4 – Extreme barrier	1– Very Bad 2– Bad 3–Good 4– Very Good 5- Excellent	1 – Not at all responsible 2 – somewhat responsible 3 – mostly responsible 4 – completely responsible	1 – much worse 2 – somewhat worse 4 – somewhat better 5 – much better	1 – not at all influential 2 – slightly influential 4 – very influential 5 – extremely influential

Source: Vagias, (2006).

As can be seen from Table 2, there are different scale points that a researcher can use, even though Likert himself used a 5-point scale. Johns (2010) argued that there is no theoretical reason to rule out different points of the response scale; after all, the options are supposed to reflect an underlying continuum rather than a finite number of possible attitudes. Perhaps the reason five is the most commonly used is because of its perfectly bipolar (-2 to +2) nature and, therefore, providing two response options to either side of the two-directional scale, is more manageable than 7, 9, or 11.

The Likert Scale Issues

It has long been established that psychometric scales, such as the Likert scale, are expected to allow survey participants to express both the direction and the strength of their opinion on a topic under consideration. The Likert scale requires the assignment of numbers and verbal labelling to rather slippery psychological phenomena such as attitude. Many researchers choose to use them with little apparent consideration of their nature, although some researchers do recognise the harmful problems inherent in such data collection techniques. For example, Peter & John (2003) are among those researchers who lament that the common practice of designing odd-numbered or midpoint scales, often with “neutral” or “undecided” midpoints, certainly removes the illusion of ordinal or interval data that the scale is expected to generate.

However, a section of the literature on the Likert scale has hailed the inclusion of the midpoint, citing the fact that it allows participants to express a neutral opinion between, for example, disagreement on one side and agreement on the other. Likert scale types without a midpoint are also characterized as forced-choice scales, as participants are forced to either choose a disagreement or agreement option (Chyung et al., 2017). On the other hand, there is a growing concern that, among other things, the midpoint may

be used as a dumping ground or an exit when participants in a survey are unfamiliar with the survey items or when the items are ambiguous or socially undesirable.

In quantitative studies, Likert scales are always associated with a continuum. Generally speaking, therefore, when survey participants are responding to a Likert scale like the one in Figure 1, they tend to have a cognitive representation of agreement dimensions ranging from “Strongly Disagree” to “Strongly Agree” with numerical values of 1 to 5, respectively.

However, in Figure 1, while the descriptive anchors extend from “Strongly Disagree” to “Strongly Agree”, which is widely considered to represent a bipolar (negative from zero to positive or ± 2) scale in a continuum, the assignment of numbers is rather unidirectional or unipolar (i.e., 1 to 5). Conspicuously, this practice of verbal and numerical labelling provides participants with contradictory stimuli because the numerical values indicate an ascending or descending order with a two-directional (bipolar) degree of agreement (Figure 1). When this happens, participants still respond to the survey anyway. It goes without saying that, in many instances, what the scales represent may be different for both researchers and participants, and the fact that how the scale is understood by participants is not the same as how researchers want the scale to be decoded, the findings of such studies are likely to mislead the world.

Furthermore, descriptive anchors are typically captured as “nominal categories” rather than “continuum,” with a midpoint labeled as “neutral,” “undecided,” or “not sure” and assigned a numerical value of 3. This “Neutral,” the “Undecided,” or the “Not Sure” breaks the supposed continuum nature of the scale, and the participants will respond to, for example, the “Neutral = 3” as higher than “Strongly Disagree” or “Disagree,” which are usually assigned with 1 or 2 respectively on a five-pointed scale, when in fact the “Neutral” should have been assigned with 0 since the assignment of the

numbers presents the scale as more of a continuum than nominal categories. In other words, the participants do not view number 3 as "Neutral," 'Undecided' or 'Not Sure', but rather as a continuation of 1 and 2. They have interpreted it as the average or the middle number, which is higher than 1 and 2 but less than 4 and 5, hence a continuum. It is noteworthy that even the original scale invented and published by Rensis Likert himself is characterized by this feature, and the same can be said of the scale used in the famous SERVQUAL study by Parasuraman et al. (1985). They used descriptive anchors that represent a categorical stream rather than a continuum. This is typical of old scales, and with this feature in a scale, research is not only likely to produce findings with a muted quality, but it is also likely to result in incorrect findings and, thus, mislead the world.

Finally, and on a lighter note, the use of the word "Neutral" or "Undecided" to express a neutral opinion on, say, an agreement scale or a satisfaction scale is unprofessional. This is because the word "Neutral" means nothing on an agreement scale and cannot be used to express a neutral stance, but what does, in the case of an agreement scale, is "neither agree nor disagree." Along the same lines, on a satisfaction scale, "neither satisfied nor dissatisfied" will do better than "Neutral" or "Undecided". Therefore, the wording of descriptive anchors should reflect their scales, as they are likely to have a lot of meaning for participants and to collect data that will lead to reliable and valid findings.

How to Properly Handle Likert Scales

As can be seen from the foregoing discourse, Likert scales, when not handled properly, can lead to the production of misleading data and findings. So, to properly handle Likert scales and eliminate or at

least reduce errors associated with their incorrect application, we proposed, as discussed below, some measures that a researcher should take at the data analysis and study design stages of a study.

The Study Design Stage

Proper handling of Likert scales is better done at the design stage of a study than at the stage of data analysis. Because at the study design stage, you can avoid all the possible mistakes you can make on the scale. Some of the steps to be taken to properly handle Likert scales at this stage are: clarifying and rewording the scales' preamble; changing some of the descriptive anchors, especially where they include "Not Sure", "Neutral" or "Undecided", and population refinement.

Population Refinement

The population of a study is defined as the total set of observations from which a sample is drawn. It is the complete set of individuals (subjects), objects, or events that have common observable characteristics that the researcher is interested in studying (Alhassan, 2020). Population refinement is the process of specifying your General Population (GP), as defined above, to reach the required sample, called the Accessible Population (AP), that has the relevant information and is available and willing to respond to your instrument (Asiamah et al., 2017).

The conceptualization of population refinement, as depicted in Figure 3, was originally designed by Asiamah et al. (2017) for a qualitative inquiry. However, it remains relevant in its quantitative counterpart, as the concept is best used with large sample sizes, which is a common feature of quantitative research.

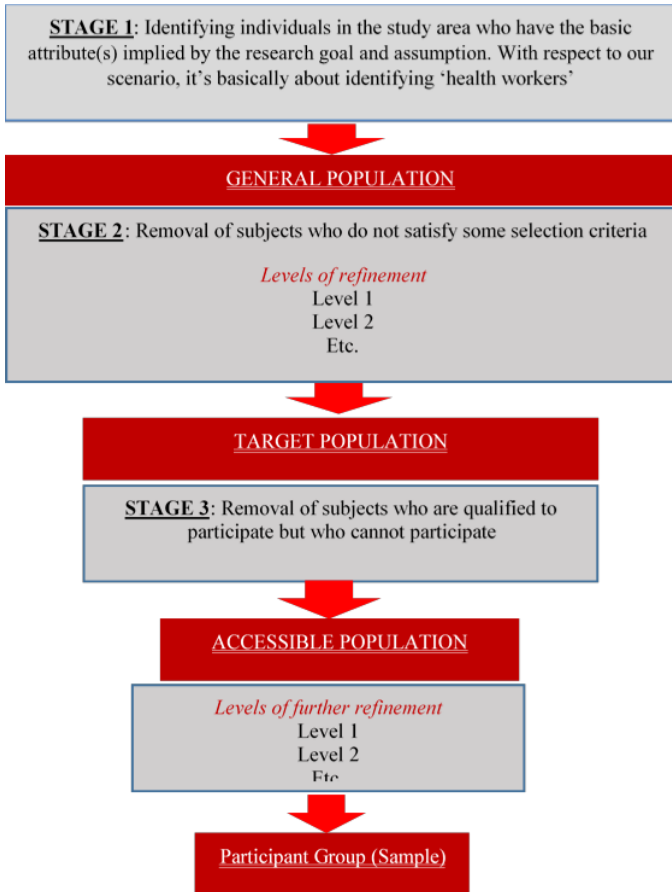


Figure 3: Steps Involved in Population Refinement

Source: Adopted from Asiamah et al., (2017).

Population refinement aims to reduce non-response rates and eliminate or, at least, reduce "Neutral," "Undecided," or "Not Sure" responses. We argue that when effective sampling is done through population refinement, then almost all survey participants will have something to say, and there may not be the need to include "Neutral," "Undecided," or "Not Sure" in the scale.

Clarifying and Rewording Scale's Preamble

A Likert scale will have to be reworded if its preamble from the source is not fully presented.

This is because the main function of the preamble is to inform the participants about the topic being investigated and what is generally expected of them. Consider the following example:

“On a scale of 1 to 5, where 1 = strongly disagree and 5 = strongly agree, indicate the extent to which you agree to the following statements”.

The above example perfectly presents the scale as though it is a continuum, yet it is not. In other words, the example hints that the scale is unipolar, which is not the case. Many researchers (Isaac, Zhiwei, & Cephas, 2022; Mohammed, Asif, Ullah, & Huma, 2021; Humphrey & Nimako, 2013; Nimako, 2013) introduce bipolar scales in this way, which is incorrect and is the basis on which most researchers communicate incorrect findings to the world. In instances like these, there is a need to reword the preamble for clarity of presentation, as done below:

“On a scale of 1 to 5, where 1 = strongly disagree, 2 = disagree, 3 = not sure, 4 = agree, and 5 = strongly agree, indicate the extent to which you agree to the following statements”

Note that for bipolar scales, capturing all the descriptive anchors in the preamble does not necessarily solve the problem. This is rightly so because a fully explained bipolar scale can be misunderstood or misinterpreted by participants. So, what you can do is change the labeling of the descriptive anchors, especially the midpoint.

Changing Descriptive Anchors

Replace “3 = not sure/neutral/undecided” with “3 = somewhat agree” to harmonize verbal labelling (descriptive anchors) with numbering/coding as this will produce a continuum-stream scale. However, bear in mind that rewording any part of a scale changes the psychometric structure of the scale and, therefore, requires robust statistical validation in data analysis. The changes will also have to be reported and justified.

The Data Analysis Stage

If, for any reason, a researcher did not properly structure the Likert scale at the design stage of the study and has, therefore, used the scale to collect data, a few techniques can be applied at this stage to at least minimize the effect of the error on the research findings. These techniques include *recoding*, *interpreting data*, and *removing a descriptive anchor*.

Removal of a Descriptive Anchor

Participants who chose “Not Sure / Neutral / Undecided” are removed from the data set. Of course, this remedy can reduce the size of the data and the sample. As such, sample size calculations should take this step into account by inflating the sample size at the sampling and data collection stage of a study. It also implies that this remedy is suitable where the data is large or where the “Not Sure/Neutral/Undecided = 3” responses are few.

Recoding

A better solution is to recode your descriptive anchors. If, for example, the original codes are:

“1 = strongly disagree, 2 = disagree, 3 = not sure, 4 = agree, and 5 = strongly agree.”

Then a researcher can recode this in data analysis as:

“0 = not sure, 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.”

This remedy also does not solve the problem because “not sure = 0” still provides a reason to worry about the recoded data since it will introduce some noise into the data. Consequently, this will be harmful to the findings of the study because the participants did not see “Not Sure = 0” but it was described to them as “Not Sure” should be equal to 3. Thus, they see it as being in the middle or average and a build-up on 1 and 2. So, if you recode “Not Sure = 0,” you are still likely to mislead the world with your findings. It only becomes a remedy if we

assume that the participants interpreted 'Not Sure' to mean a state of ignorance or neutral stance, not the average or the midpoint.

Data Interpretation

You can present your findings in such a way that your Likert scale makes sense during the data interpretation stage. If you are unable to take these steps, then you may not be able to treat variables measured with a Likert scale as continuous or ordinal. Because in quantitative research, Likert scales are transformed into continuous variables. The transformation is usually done by parcelling, where all the ratings are aggregated into a score called an index, which provides a composite score for a trait of interest.

So, if a Likert scale cannot be transformed into continuous or ordinal data, then a researcher may not be able to perform correlational analyses on such scales. If you do, you will be misleading the world because a positive correlation could turn out to be negative, and vice versa. In addition, an insignificant relationship could become significant and vice versa, leading to Type I and II errors. In this sense, flawed key descriptive measures can be obtained simply by using '3 = Neutral / Undecided / Not Sure' instead of '0 = Neutral / Undecided / Not Sure'. When you communicate such findings to the world, people can make decisions based on them, and this could be very dangerous to science and humanity. As a result, the only thing a researcher can do with categorical Likert scale data is perform descriptive statistics. Even under these circumstances, the researcher will have to emphasize the neutral role of "Not Sure" so that participants know that "Not Sure" represents a state of ignorance or a neutral posture.

Proposed Best Practice

Use a validated unipolar scale with anchors that represent a continuum: this is the best remedy for the problem discussed in the previous sections.

DOI: <https://doi.org/10.47740/621.UDSIJD6i>

Because both the labelling (descriptive anchors) and the numerical assignment will be harmonious if this advice is heeded. For example, you can have

the descriptive anchors and their numerical values harmonized as shown in Figure 4:

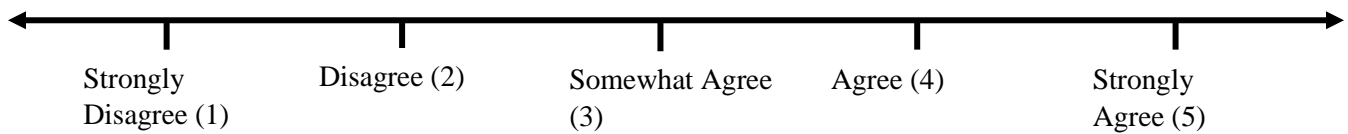


Figure 4: Proposed Harmonized Likert Scale

Source: Authors' Elaboration, 2021.

In this example, both the descriptive anchors and the numbering are in ascending (or decreasing) order, making it a continuum.

Preambles should clearly explain the role of each descriptive anchor and its score: inform participants that if they choose 3, they are saying they know nothing. That is, they are not sure, are neutral, or are undecided about the statement. This can be done by providing a proper explanation for each descriptive anchor and its score.

Use descriptive anchors without scores: especially if you are using a 3-point or a bipolar Likert scale. Sometimes, because numbering can confuse most participants, it will be better to use only descriptive anchors so that you can code the responses after data collection. For example, you can use “Neither Agree nor Disagree, Strongly Disagree, Disagree, Agree, Strongly Agree” without assigning numbers to them. By doing this, participants are likely to choose responses that truly reflect their opinions.

Modification of Descriptive Anchor Before or After Data Collection: You can also modify descriptive anchors before or after data collection, as this will allow you to reconcile categorical labelling with the numbering or coding, thereby making the scale more of a continuum.

Conclusion and Implication for Future Research

Our main objective was to point out the mishaps relating to the inappropriate use of the Likert scale.

Specifically, we have discussed how the midpoint of a Likert scale disrupts or discontinues the supposed continuum nature of a typical Likert scale, especially of the bipolar type. Furthermore, we have suggested some practical remedies for reversing errors that might occur as a result of the inappropriate use of the Likert scale, and finally, we proposed a harmonized Likert scale that takes care of the conflicting functions of descriptive anchors and numerical labelling of the scale.

To this end, researchers are expected to thoroughly examine the objectives of their studies, as this will largely inform their choice of statistical data analysis techniques. Even so, the researcher's knowledge of the types of data produced by Likert scales must be deployed to adapt the most complete and robust framework of what is expected of such data.

References

- Adelson, J. L., & McCoach, D. B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point likert-type scale. *Educational and Psychological Measurement, 70*(5), 796–807. <https://doi.org/10.1177/0013164410366694>
- Alhassan, I. (2020). *An assessment of the operationalisation of Ghana's policy on free senior high school education in some selected public senior high schools in the Tamale Metropolis,*

- Tamale* [Unpublished Master of Philosophy Thesis], University for Development Studies.
- Asiamah, N., Mensah, H. K., & Oteng-Abayie, E. F. (2017). General, target, and accessible population: Demystifying the concepts for effective sampling. *Qualitative Report*, 22(6), 1607–1621. <https://doi.org/10.46743/2160-3715/2017.2674>
- Bhandari, P. (2020, October 12). *Designing and analyzing a Likert Scale what, why and how*. <https://www.scribbr.com/methodology/likert-scale/>
- Bhandari, P. (2022, July 3). *Designing and analyzing a Likert scale: Guide & examples*. <https://www.scribbr.com/methodology/likert-scale/>
- Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data. *The Journal of Extension*, 50. <https://www.semanticscholar.org/paper/Analyzing-Likert-Data.-Boone-Boone/84df2fd8bd96bf28009fb7f8d035dfef63b3b8dd>
- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18(3), 205–215. <https://doi.org/10.1177/014662169401800302>
- Chyung, S. Y. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, 56(10), 15–23. <https://doi.org/10.1002/pfi.21727>
- Durand, R., M., & Lambert, Z., V. (1988). Don't know responses in surveys: Analyses and interpretational consequences. *Journal of Business Research*, 16(2), 169–188. [https://doi.org/10.1016/0148-2963\(88\)90040-9](https://doi.org/10.1016/0148-2963(88)90040-9)
- Guy, R. F., & Norvell, M. (1977). The neutral point on a Likert Scale. *The Journal of Psychology*, 95(2), 199–204. <https://doi.org/10.1080/00223980.1977.9915880>
- Humphrey, D., & Nimako, G. S. (2013, April). Examining the Undergraduate Construction Technology Students' Use of Computer and Internet in UEW: A Structural Equation Modelling Approach. *International Journal of Scientific & Technology Research*, 2(4), 90-98.
- Isaac, E. D., Zhiwei, T., & Cephas, P. K. (2022, June 16). Usage Intention of e-Learning Systems in Ghanaian Tertiary Institutions: A Case Study of the University for Development Studies. *Sustainability*, 1-18.
- Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, 7(4), 396–403. <https://doi.org/10.9734/bjast/2015/14975>
- Jovancic, N. (2020, January 17). *What is the Guttman scale and how to use it in your surveys*. <https://www.leadquizzes.com/blog/guttman-scale/>
- Kulas, J., Stachowski, A., & Haynes, B. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business and Psychology*, 22, 251–259. <https://doi.org/10.1007/s10869-008-9064-2>
- 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). *A technique for the measurement of attitudes* [Doctor of Philosophy Thesis], Columbia University. <https://www.worldcat.org/title/technique-for-the-measurement-of-attitudes/oclc/812060>
- Mohammed, A., Asif, A. S., Ullah, A., & Huma, I. (2021, February). A Study of Impacting Factors on Technology Adoption in the Public Sector of Pakistan. *Journal of Contemporary Issues in Business and Government*, 27, 1281-1302.
- Nimako, S. G. (2013, August). Critical ATM Adoption Factors in Ghanaian Banking Industry: CFA

Approach. *Research Journal of Social Science & Management*, 3, 91-102.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *The Journal of Marketing*, 49(4), 41-50. <http://www.jstor.org/stable/1251430?origin=JSTOR-pdf>

Peter, F. (2018, August 14). *Nominal, ordinal, interval, ratio: Stevens' typology and some problems with it*. <https://medium.com/peter-flom-the-blog/nominal-ordinal-interval-ratio-stevens-typology-and-some-problems-with-it-93b1033815f7>

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(1)1–15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)

Raaijmakers, Q. A. W., Hoof, J. T. C., Hart, H., Verbogt, T. F. M. A., & Vollebergh, W. A. M. (2000). Adolescents' midpoint responses on Likert-

type scale items: Neutral or missing values? *International Journal of Public Opinion Research*, 12. <https://doi.org/10.1093/ijpor/12.2.209>

Rinker, T. (2014). Likert - Running head: The treatment of Likert data. <https://www.coursehero.com/file/10349721/Likert/>

Rob, J. (2010, March 3). *Likert Items and Scales*. <https://pdfslide.net/documents/likert-items-and-scales-the-university-of-sheffield-filelikertfact1-survey.html>

Roufun, N. (2018, November). *Adaptation of psychometric tools in Bangladesh*. <https://www.ukessays.com/essays/psychology/adaptation-psychometric-tools-8255.php?vref=1#citethis>

Stevens, S. S. (1946). On the theory of scales of measurement. *New Series*, 103(2684), 677–680. <http://www.jstor.org/stable/1671815>

Vagias, W. M. (2006). *Likert-type scale response anchors*. Clemson International Institute for Tourism & Research Development.