encyclopedia

MDPI

Entry

Likert-Type Scale

Malcolm Koo [1,2,*] **and Shih-Wei Yang** [3,4]

[1] Department of Nursing, Tzu Chi University, Hualien 970302, Taiwan
[2] Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada
[3] Department of Education, National Chiayi University, Minsyong Township, Chiayi 621302, Taiwan
[4] Department of Styling and Cosmetology, Tainan University of Technology, Tainan City 710302, Taiwan
* Correspondence: m.koo@utoronto.ca

Definition: The Likert-type scale is a widely used psychometric instrument for measuring attitudes, opinions, or perceptions in research contexts. It presents respondents with a series of statements accompanied by symmetrical response options, typically structured on a five-point scale ranging from "Strongly Disagree" to "Strongly Agree". Each point on the scale represents a gradation of agreement or sentiment, allowing researchers to transform subjective responses into quantifiable data for statistical analysis and interpretation.

Keywords: Likert scale; measurement; psychometrics; scale development; questionnaire design; survey design

1. Introduction

1.1. Overview of the Likert Scale

The Likert scale is a widely utilized rating scale for measuring unobservable constructs, such as attitudes, opinions, or perceptions through a structured response format [1]. It consists of a series of statements or questions, each accompanied by a symmetric range of response options that allow respondents to indicate their level of agreement or disagreement. A typical five-point scale includes options ranging from "Strongly Disagree" to "Strongly Agree", with intermediate responses such as "Disagree", a neutral "Neither Agree nor Disagree", and "Agree". The term "Likert-type scale" encompasses both the original methodology introduced by Rensis Likert and its variations, including scales with different numbers of points (e.g., 4 to 7 points).

The Likert-type scale is a valuable research tool due to its ability to capture human sentiments in a standardized manner, enabling respondents to express varying degrees of opinion rather than binary choices. By transforming subjective qualitative data into quantifiable metrics, it facilitates detailed data collection and robust statistical analysis. Furthermore, aggregating responses across multiple items offers a comprehensive measure of underlying attitudes or opinions. This versatility makes the scale an indispensable tool in disciplines such as psychology, marketing, social sciences, and healthcare, where understanding participants' perspectives is critical.

1.2. Historical Background and Development

The Likert (pronounced as "Lick-urt") scale was developed by American social scientist Rensis Likert (1903–1981) in 1932 as part of his doctoral thesis, *A Technique for the Measurement of Attitudes* [2]. In subsequent work, Likert and his colleagues refined the process by simplifying the Thurstone scaling technique, a contemporary method for measuring latent variables such as attitudes on a continuous scale.

check for updates

Academic Editors: Kamran Munir and Elena-Mădălina Vătămănescu

Received: 9 December 2024
Revised: 15 January 2025
Accepted: 24 January 2025
Published: 3 February 2025

Citation: Koo, M.; Yang, S.-W. Likert-Type Scale. *Encyclopedia* **2025**, *5*, 18. https://doi.org/10.3390/encyclopedia5010018

The Thurstone method, one of the earliest techniques for measuring latent variables such as attitudes on a continuous scale, required experts to evaluate and assign scale values to statements, making it resource-intensive and dependent on complex scoring procedures. The Likert scale addressed these challenges by introducing a simplified scoring system, where numerical values were directly assigned to participant responses. This approach made the process more efficient, consistent, and reliable, representing a significant advancement over the original Thurstone method [3].

Over the decades, the original scale has undergone numerous adaptations to meet diverse research needs. These include even-numbered scales, which eliminate the neutral midpoint and compel respondents to take a definitive stance, as well as seven-point scales that offer greater response granularity. Moreover, the scale's language and response options have been customized to align with specific research contexts, shifting from the traditional agreement-based format to alternatives that measure frequency, importance, satisfaction, and other perceptual dimensions. Despite its widespread adoption, debates continue regarding the appropriate analysis and interpretation of Likert-type scales, particularly whether their response options should be treated as ordinal or interval data [4].

2. Design of the Likert-Type Scale

2.1. Likert Items and Likert-Type Scales

The terms Likert item and Likert-type scale are *often* used interchangeably but represent distinct concepts in survey research. A Likert item refers to a single question with a Likert-type response format, while a Likert-type scale is a composite measure comprising several (typically 4 to 10) related Likert items designed to assess a broader construct. Responses to individual items are summed or averaged to generate an overall score representing the respondent's position on the construct (Table 1).

Table 1. Comparison of Likert items and Likert-type scales.

Aspect	Likert Items	Likert-Type Scales
Scope	Measures a single, specific attitude or perception	Measures an overall construct or latent variable
Composition	Consists of one question with ordinal responses	Composed of multiple, typically 4 to 10, related Likert items
Scoring	Response analyzed individually	Responses summed or averaged across items

An example of a Likert-type scale is the 36-item Short Form Health Survey (SF-36) questionnaire, a standardized tool for measuring health-related quality of life across eight domains of health. The SF-36 uses individual Likert items, with their scores aggregated into two composite scores representing physical and mental health [5].

2.2. Formulating Likert-Type Scale Questions

Designing effective Likert-type scale questions is essential for capturing accurate and meaningful data in survey research. The quality of the data collected depends on the clarity and precision of the questions posed. As with general principles of questionnaire design [6–8], each Likert-type item must be carefully constructed to elicit reliable and valid responses that truly reflect the respondents' attitudes or opinions.

Questions should be clearly worded, concise, and devoid of ambiguity to ensure consistent interpretation among respondents. Ambiguous or complex phrasing can lead to misinterpretation, resulting in data that do not accurately represent the respondents' true feelings. Each item should address a single idea or statement to minimize confusion

and facilitate precise responses. Double-barreled questions—those that ask about two issues at once—should be avoided as respondents may agree with one part of the question while disagreeing with the other. Similarly, double-negatives should also be avoided as they increase the cognitive burden and the likelihood of misreading or misunderstanding the statement. In addition, items should be phrased neutrally to avoid biasing responses. Leading or loaded terms that suggest socially desirable answers can skew results and undermine the objectivity of the data.

Framing Likert-scale items in an interrogative format, rather than an assertive format, has been suggested to encourage respondents to engage more critically with their answers, thereby reducing response biases [9,10], such as default response bias and acquiescence bias. Default response bias occurs when individuals select options based on ease or accessibility, such as the first choice, the midpoint, or other readily available responses. Acquiescence bias, the tendency to agree with statements regardless of their content, introduces systematic errors that can distort survey outcomes.

To mitigate these issues, it is essential to pre-test Likert-scale items with a small, representative sample before full deployment. Pre-testing helps identify potential problems with wording or interpretation, allowing researchers to refine the questions for improved clarity and consistency. Feedback from the pre-test ensures that the items are well understood and appropriately capture the intended constructs, enhancing the overall reliability and validity of the survey.

2.3. Choice of Likert Item Options

Choosing effective Likert item options is crucial for capturing accurate and meaningful data in survey research. A well-designed Likert item provides a clear, balanced range of response options that accurately reflect the respondents' attitudes or opinions. In addition, including a brief instruction or preamble at the beginning of the Likert-type scale is essential to guide respondents on how to interpret and use the scale correctly. For example, a preamble might state: "On a scale of 1 to 5, where 1 = Strongly Disagree and 5 = Strongly Agree, please indicate the extent to which you agree with the following statements". This introductory instruction ensures that respondents understand the scaling system and the meaning of each response option.

Likert item options should typically range from four to seven points to provide a balanced and symmetrical scale of choices [11]. A balanced scale ensures that respondents have an equal opportunity to express positive and negative sentiments, as well as neutrality, if applicable. This balance minimizes response bias and enhances the reliability of the data collected.

For example, in a common five-point Likert-type scale, the following response anchors may be used:

- Strongly Disagree;
- Disagree;
- Neither Agree nor Disagree;
- Agree;
- Strongly Agree.

This configuration includes two negative options, a neutral midpoint, and two positive options, providing a symmetrical range around the neutral point.

Deciding the number of response options involves balancing the need for detailed data with the risk of respondent fatigue. A greater number of points enables finer distinctions in respondents' attitudes, enhancing the sensitivity of the measurement [12]. For instance, a seven-point scale might include additional options such as "Slightly Disagree" and "Slightly Agree", allowing for subtler differences in opinion.

However, too many response points can overwhelm respondents, potentially leading to respondent fatigue. This fatigue can manifest as reduced attention, rushed responses, or even survey abandonment, all of which can compromise data quality [13]. A review study of 60 articles using Likert scales, published between 2012 and 2021, concluded that odd-numbered response scales of more than five points, especially seven-point scales, were the most effective in terms of reliability and validity [14].

Another study that compared data characteristics of Likert-type scales with 5-point, 7-point, and 10-point response formats found that the 5- and 7-point scales produced the same mean scores when rescaled, whereas the 10-point format tended to produce slightly lower relative means. Notably, the scale format had minimal impact on other data characteristics, such as variation around the mean, skewness, or kurtosis. These findings suggest that while the number of points on the scale can influence mean scores, other statistical properties remain largely consistent across formats [15].

The decision to include a neutral response option should be guided by the research objectives. A neutral option, such as "Neither Agree nor Disagree", allows respondents to avoid forced choices that might not accurately represent their true perspective. However, some researchers argue that neutral options can attract indecisive respondents or those unwilling to take a stance. This disengagement may stem from factors such as survey fatigue, lack of comprehension, time pressure, or a reluctance to commit. Consequently, the inclusion of a neutral option can dilute the discriminative power of the data. Researchers must carefully weigh these trade-offs and consider whether a forced-choice format with an even number of response options better aligns their study objectives [16]. Pilot testing can help determine the most effective approach for the target population and research question.

Using terms such as "Neutral" or "Undecided" as response options in agreement or satisfaction scales is generally discouraged, as these terms may not accurately reflect a neutral stance within the specific context of the scale. For example, on an agreement scale, "Neutral" does not clearly correspond to the continuum between agreement and disagreement. Instead, using "Neither Agree nor Disagree" more explicitly conveys a middle-ground position on the statement [14].

The visual layout of Likert responses, particularly in online surveys, can influence respondents' cognitive and perceptual processes. Vertical Likert-scale formats tend to elicit higher rates of extreme responses compared to horizontal formats. This difference arises because the compact visual arrangement of vertical layouts makes endpoints appear less extreme and more accessible. In contrast, horizontal formats span a wider visual range, requiring greater effort to select extreme options, thereby reducing extreme response tendencies [17].

In addition, the choice of graphical elements, such as radio buttons versus square checkboxes, may affect user perception and response accuracy. Radio buttons, commonly associated with single, mutually exclusive choices, are the standard graphical element for Likert-scale items. Square checkboxes, on the other hand, suggest the possibility of multiple selections, potentially introducing cognitive ambiguity and reducing the clarity of the measurement instrument [18].

If numerical labels are used alongside text anchors in the Likert response set, careful consideration is crucial to minimize response bias and enhance data quality. The selection of numerical values can significantly influence participants' interpretation and response patterns. For example, bipolar scales (e.g., −2, −1, 0, 1, 2), rather than unipolar (1, 2, 3, 4, 5), may unconsciously discourage respondents from selecting extreme negative values due to leniency or positivity bias [19].

Moreover, a well-designed Likert response set should include a clear midpoint (e.g., 3 on a 1–5 scale) to accommodate respondents with neutral or uncertain attitudes. A 1–5 scale,

rather than alternatives such as a 0–4 scale, offers the advantage of an intuitively clear arithmetic midpoint. In contrast, the 0–4 scale lacks a readily recognizable central value, which may lead to confusion or inconsistent responses. Furthermore, using zero as an anchor in scales can be problematic as it often represents a cognitively absolute value, frequently interpreted as a symbol of extremity or absence. This interpretation may create a psychological barrier, discouraging respondents from selecting it.

2.4. Cross-Cultural Considerations in Likert-Type Scales

Cultural considerations should be kept in mind when using Likert-type scales [20]. Cultural differences can influence how respondents interact with Likert response formats. For instance, some cultures may exhibit a preference for extreme responses (e.g., Strongly Agree or Strongly Disagree), while others may gravitate toward neutral options [21]. The wording of the questions must also be carefully translated to preserve their meaning and intent across languages. Furthermore, the content of the statements must be culturally relevant and sensitive, as questions appropriate in one culture may be misunderstood or even deemed offensive in another.

In some cultures, the respondents may have a greater tendency to provide socially desirable answers rather than their true opinions, particularly when addressing sensitive topics [22]. To address these challenges, it is often necessary to conduct pilot testing with the target population. Pilot tests can help ensure that Likert items are understood as intended and that the scale's design is appropriate for the specific cultural context.

3. Data Analysis Techniques

3.1. Coding Likert-Item Responses

Coding Likert-item responses is a critical step in converting qualitative survey data into quantitative data for statistical analysis. Each response option on a Likert-type scale is assigned a numerical value, enabling researchers to perform mathematical operations and draw meaningful interpretations from the data. Typically, coding is carried out sequentially, with the most negative response assigned the lowest value and the most positive response assigned the highest value. For instance, in a five-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree", the coding might follow this scheme:

Strongly Disagree—1;
Disagree—2;
Neither Agree nor Disagree—3;
Agree—4;
Strongly Agree—5.

When multiple Likert items are designed to assess a single construct, their coded responses can be summed or averaged to create a composite score. This composite score provides a more reliable and valid measure of the construct than any single item as it reduces the impact of random errors associated with individual items. Combining items assumes that all items are equally weighted and contribute similarly to the overall construct. The process for constructing summative Likert-type scales includes several key steps: defining the construct, developing an item pool, selecting the number of scale steps, choosing appropriate anchors, piloting the scale, and calculating reliability and validity, as detailed by Viljoen [23].

The use of positively and negatively worded Likert items has been widely debated in survey methodology. Advocates for mixing item types argue that including both positively and negatively worded statements can counteract acquiescence bias. By reversing the wording of certain items, the respondents are encouraged to read each item more carefully, thus providing more thoughtful and accurate answers [24].

However, empirical evidence suggests that this approach may inadvertently impair response accuracy due to increased cognitive complexity. Negatively worded or reverse-coded items require additional mental effort to interpret, which can lead to confusion, misreading, or misunderstanding. This, in turn, may result in inconsistent or erroneous responses, undermining the reliability of the data [25,26].

A study that analyzed data from 700 respondents using the Multidimensional Fatigue Inventory compared the psychometric properties of scales containing uniformly worded items with those including both straightforward and reverse-worded items. The findings revealed that reverse-worded items were ineffective in mitigating response bias. Instead, they increased the likelihood of errors due to respondent inattention and confusion. Based on these results, the study recommended using items consistently worded in the same direction, particularly in epidemiological and clinical studies, to improve data quality and reliability [27].

Nevertheless, if reverse-worded items are indeed used in a composite Likert-type scale, their coding must be reversed during the scoring process. This step ensures consistency in the direction of coding across all items, allowing for an accurate calculation of composite scores.

3.2. Analyzing Likert-Type Data

Analyzing Likert-type data is a critical step of survey research as it directly impacts the validity of the conclusions drawn. A key debate revolves around whether Likert-type data should be treated as ordinal- or interval-level (continuous) data [28–30] (Table 2). Jamieson argued that non-parametric tests are more appropriate for analyzing Likert-type data as parametric tests assume interval- or ratio-level data [4]. In contrast, Norman [31] asserted that parametric statistics can be applied to Likert data, even under conditions such as small sample sizes, unequal variances, or non-normal distributions. Moreover, it has been demonstrated that the F-test used in the analysis of variance (ANOVA) is robust to violations of the interval data assumption, with no significant bias observed when data are collected using a 5- to 7-point Likert response format [32].

Table 2. Data analysis techniques for Likert-type scale data.

	Individual Likert Items	**Composite Scales**
Descriptive statistics	Median, mode, frequency, and percentage distributions	Mean, standard deviation, Cronbach's alpha, McDonald's omega
Comparative tests	Mann–Whitney U test (for 2 groups) Kruskal–Wallis test (for $\geq$2 groups) Chi-square test for distribution comparisons	*t*-tests analysis of variance regression analysis
Correlation analysis	Spearman's rank-order correlation	Pearson's correlation
Advanced techniques	Not applicable	Factor analysis

When multiple Likert items are summed or averaged to create a composite score, the resulting data can often be treated as approximately continuous. This allows for the use of parametric tests, such as t-test, ANOVA, Pearson's correlation coefficients, and multiple regression analyses [33]. Furthermore, advanced statistical techniques, such as factor analysis, can be employed to identify underlying dimensions within the data. Reliability analysis methods, such as Cronbach's alpha or MacDonald's omega, can be used to assess the internal consistency of the scale [34].

In contrast, non-parametric statistics, such as the Mann–Whitney U test or Kruskal–Wallis test, are generally recommended for analyzing single Likert items, as these items are ordinal data where the assumption of equal intervals is often questionable [35]. Neverthe-

less, results from simulations have shown that for comparing two samples of 5-point Likert items, the t-test and Mann–Whitney U test yield similar statistical power, with type I error rates for both methods being less than 3% above the nominal rate of 5% [36].

For single Likert items, it is generally more appropriate to use the median or mode rather than the mean as a measure of central tendency. The mean assumes equal intervals between scale points, which may lead to misleading conclusions if the data are skewed or contain outliers. Moreover, presenting the frequency and percentage of each response option can offer valuable insights into the distribution of responses. For instance, while a mean or median of 3 on a 5-point Likert scale might suggest neutrality, this single value does not reveal the shape of the distribution. A mean of 3 could represent scenarios where half the respondents chose "Strongly Disagree (1)" and the other half "Strongly Agree (5)", or where all respondents selected "Neither Agree nor Disagree (3)". Despite yielding the same mean, these scenarios have distinctly different implications.

To simplify analyses, researchers often collapse multiple response categories into two or three broader categories. This approach enhances interpretability, particularly for smaller sample sizes or when some categories are sparsely populated. For example, a 5-point Likert scale might be collapsed into two categories, such as "Agreement" (combining "Strongly Agree" and "Agree") versus "Non-agreement" (combining "Neutral", "Disagree", and "Strongly Disagree"). Alternatively, three categories could be retained, such as "Agreement", "Neutral", and "Disagreement". Collapsing categories helps to focus the analysis on broader trends and satisfies assumptions of statistical tests, such as chi-square or logistic regression. When collapsing categories, researchers must clearly document and justify the criteria used. This transparency is essential to ensure the validity and interpretability of the results while maintaining the integrity of the original data.

3.3. Visualization of Likert Data

In addition to tables, graphical methods are invaluable for presenting response distributions and identifying data trends and patterns. Given that individual Likert items represent ordinal data, visual presentations should follow their categorical nature. A straightforward approach is to use bar charts to display the frequency or percentage of responses for each category of a Likert item. This method provides an intuitive way for readers to quickly assess the distribution of responses across all categories [37].

For comparing responses across multiple Likert items within a single visual, horizontal stacked bar charts are particularly effective. Each bar represents a different Likert item, with segments within the bar denoting the proportion of responses in each category. By normalizing them to 100%, the focus shifts to relative proportions rather than raw counts, facilitating easy comparisons between items. A horizontal orientation further enhances readability, especially when labels are included.

A more sophisticated option is the diverging stacked bar chart, which is particularly suited for visualizing polarized responses. In this format, neutral responses are centered, with agreement and disagreement categories diverging to the right and left, respectively. This design effectively illustrates the balance between positive and negative sentiments. Aligning the neutral category at zero creates symmetry, making the data easier to interpret at a glance. To improve clarity and readability, contrasting colors can be used for agreement and disagreement categories, or a gradient transition can represent varying levels of intensity.

While basic bar charts and stacked bar charts can be easily created using commercial software packages, such as Microsoft Excel, more advanced visuals, such as diverging stacked bar charts, are more efficiently generated using programming languages, such as

Python or R [38]. These tools offer greater flexibility and automation, enabling the creation of clear and informative visualizations that enhance the interpretation of research findings.

In R, the "likert" package offers a streamlined approach for analyzing and visualizing Likert-type data. This package simplifies the creation of diverging stacked bar charts and other graphical representations tailored to Likert scales [39]. Similarly, Python, with libraries such as matplotlib and seaborn, allows users to create custom visualizations. For example, Supplementary File S1 contains a Python script illustrating how to create a clustered diverging stacked bar chart for a 5-point Likert-type scale consisting of five items with a summation score (Figure 1).

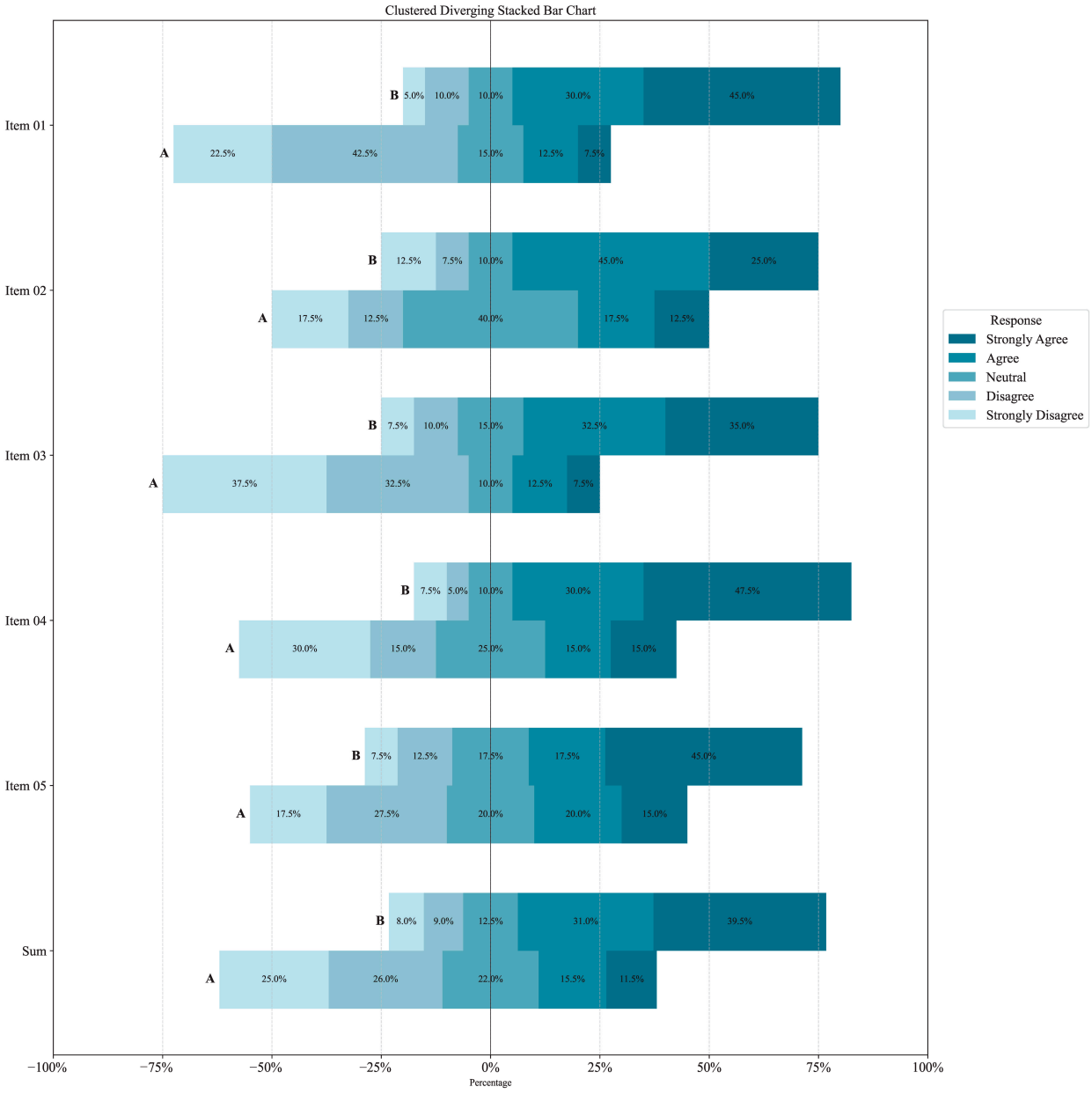

Figure 1. A clustered diverging stacked bar chart of summative 5-point Likert scale responses created with fictitious data. The chart was created using Python code, detailed in Supplementary File S1.

3.4. Handling of Missing Data

Careful consideration is essential when handling missing data in the coding and scoring process. Decisions about imputing missing responses or excluding incomplete surveys can significantly impact the results and interpretations of a study, and such decisions

must be consistent and transparent. Missing data are particularly important in summative Likert-type scales, where omitted items can lead to inaccurately depressed summed scores, resulting in invalid conclusions. To preserve the integrity of data analysis, imputation may be necessary to address missing responses.

The validity of simple imputation methods, such as substituting the group mean or the individual's mean scores from completed items, relies on the assumption that the missing data are *missing at random* (MAR). MAR implies that the probability of a response being missing is unrelated to the unobserved data and only related to the observed data. If the data are *missing not at random* (MNAR)—meaning the likelihood of missingness is directly related to the unobserved value itself—simple imputation can introduce bias. In such cases, more advanced techniques, such as multiple imputation or model-based methods, may be necessary to handle missing data appropriately [40].

Furthermore, sensitivity analyses should be conducted to evaluate the impact of missing data and imputation methods on a study's findings [41]. These analyses help to determine whether a study's conclusions are robust to various assumptions about the nature of the missing data and the techniques used to address it.

4. Conclusions

The Likert-type scale is an indispensable tool in research, offering a versatile and standardized means of quantifying subjective attitudes, opinions, and perceptions. Its straightforward format facilitates ease of administration and comprehension across diverse populations.

Despite its advantages, the Likert-type scale has notable limitations. Common challenges include response bias, such as central tendency bias, where respondents avoid extreme categories, and acquiescence bias, where individuals are inclined to agree with statements regardless of their content. Moreover, ongoing debates about whether Likert scale data should be treated as ordinal or interval can influence the choice of statistical methods and the interpretation of results.

By recognizing and addressing these strengths and limitations, researchers can make informed decisions about the appropriate use of the Likert-type scale, ensuring it aligns with their research objectives and the characteristics of their target population.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/encyclopedia5010018/s1. File S1. The Python code for creating a clustered diverging stacked bar chart (Figure 1).

Author Contributions: Conceptualization, M.K. and S.-W.Y.; writing—original draft preparation, M.K. and S.-W.Y.; writing—review and editing, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Therefore, data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Joshi, A.; Kale, S.; Chandel, S.; Pal, D.K. Likert scale: Explored and explained. *Br. J. Appl. Sci. Technol.* **2015**, *7*, 396–403. [CrossRef]
2. Likert, R. A Technique for the Measurement of Attitudes. Ph.D. Thesis, New York University, New York, NY, USA, 1932.

3. Likert, R.; Roslow, S.; Murphy, G. A simple and reliable method of scoring the Thurstone attitude scales. *J. Soc. Psychol.* **1934**, *5*, 228–238. [CrossRef]
4. Jamieson, S. Likert scales: How to (ab)use them. *Med. Edu.* **2004**, *38*, 1212–1218. [CrossRef]
5. Ware, J.E., Jr. SF-36 health survey update. *Spine* **2000**, *25*, 3130–3139. [CrossRef] [PubMed]
6. Meadows, K.A. So you want to do research? 5: Questionnaire design. *Br. J. Community Nurs.* **2003**, *8*, 562–570. [CrossRef] [PubMed]
7. Fowler, F.J., Jr.; Cosenza, C. Design and evaluation of survey questions. In *The SAGE Handbook of Applied Social Research Methods*, 2nd ed.; Bickman, L., Rog, D.J., Eds.; Sage: Thousand Oaks, CA, USA, 2009; pp. 375–412.
8. Krosnick, J.A.; Presser, S. Question and questionnaire design. In *Handbook of Survey Research*, 2nd ed.; Marsden, P.V., Wright, J.D., Eds.; Emerald Group: Bingley, UK, 2010; pp. 263–313.
9. Petty, R.E.; Rennier, G.A.; Cacioppo, J.T. Assertion versus interrogation format in opinion surveys questions enhance thoughtful responding. *Public Opin. Q.* **1987**, *51*, 481–494. [CrossRef]
10. Höhne, J.K. Question Format, Response Effort, and Response Quality: A Methodological Comparison of Agree/Disagree and Item-Specific Questions. Ph.D. Thesis, Georg-August Universität, Göttingen, Germany, 2017.
11. Lozano, L.M.; García-Cueto, E.; Muñiz, J. Effect of the number of response categories on the reliability and validity of rating scales. *Methodology* **2008**, *4*, 73–79. [CrossRef]
12. Beckstead, J.W. On measurements and their quality. Paper 4: Verbal anchors and the number of response options in rating scales. *Int. J. Nurs. Stud.* **2014**, *51*, 807–814. [CrossRef]
13. Streiner, D.L.; Norman, G.R.; Cairney, J. Scaling responses. In *Health Measurement Scales. A Practical Guide to Their Development and Use*, 5th ed.; Oxford University Press: Oxford, UK, 2015; pp. 38–73.
14. Kusmaryono, I.; Wijayanti, D.; Maharani, H.R. Number of response options, reliability, validity, and potential bias in the use of the Likert scale education and social science research: A literature review. *Int. J. Educ. Methodol.* **2022**, *8*, 625–637. [CrossRef]
15. Dawes, J. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *Int. J. Mark. Res.* **2008**, *50*, 61–104. [CrossRef]
16. Froman, R.D. The ins and outs of self-report response options and scales. *Res. Nurs. Health* **2014**, *37*, 447–451. [CrossRef]
17. Weijters, B.; Millet, K.; Cabooter, E. Extremity in horizontal and vertical Likert scale format responses. Some evidence on how visual distance between response categories influences extreme responding. *Int. J. Res. Mark.* **2021**, *38*, 85–103. [CrossRef]
18. Tourangeau, R.; Conrad, F.G.; Couper, M.P. Introduction to measurement and design in Web surveys. In *The Science of Web Surveys*; Oxford University: New York, NY, USA, 2013; pp. 57–76.
19. Armitage, C.; Deeprose, C. Changing student evaluations by means of the numeric values of rating scales. *Psychol. Learn. Teach.* **2004**, *3*, 122–125. [CrossRef]
20. Smith, T.W. Surveying across nations and cultures. In *Handbook of Survey Research*, 2nd ed.; Marsden, P.V., Wright, J.D., Eds.; Emerald Group: Bingley, UK, 2010; pp. 733–763.
21. Chen, C.; Lee, S.Y.; Stevenson, H.W. Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychol. Sci.* **1995**, *6*, 170–175. [CrossRef]
22. Harzing, A.W. Response styles in cross-national survey research. *Int. J. Cross Cult. Manag.* **2006**, *6*, 243–266. [CrossRef]
23. Viljoen, M. Constructing homogeneous Likert-type summative rating scales according to classical measurement theory. *J. Soc. Sci.* **2015**, *43*, 143–151. [CrossRef]
24. Guyatt, G.H.; Cook, D.J.; King, D.; Norman, G.R.; Kane, S.L.; Van Ineveld, C. Effect of the framing of questionnaire items regarding satisfaction with training on residents' responses. *Acad. Med.* **1999**, *74*, 192–194. [CrossRef]
25. Swain, S.D.; Weathers, D.; Niedrich, R.W. Assessing three sources of misresponse to reversed Likert items. *J. Mark. Res.* **2008**, *45*, 116–131. [CrossRef]
26. Weems, G.H.; Onwuegbuzie, A.J.; Collins, K.M. The role of reading comprehension in responses to positively and negatively worded items on rating scales. *Eval. Res. Edu.* **2006**, *19*, 3–20. [CrossRef]
27. van Sonderen, E.; Sanderman, R.; Coyne, J.C. Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE* **2013**, *8*, e68967. [CrossRef] [PubMed]
28. Harpe, S.E. How to analyze Likert and other rating scale data. *Curr. Pharm. Teach. Learn.* **2015**, *7*, 836–850. [CrossRef]
29. Sullivan, G.M.; Artino, A.R., Jr. Analyzing and interpreting data from Likert-type scales. *J. Grad. Med. Educ.* **2013**, *5*, 541–542. [CrossRef] [PubMed]
30. Bishop, P.A.; Herron, R.L. Use and misuse of the Likert item responses and other ordinal measures. *Int. J. Exerc. Sci.* **2015**, *8*, 297–302. [CrossRef] [PubMed]
31. Norman, G. Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ. Theory Pract.* **2010**, *15*, 625–632. [CrossRef]
32. Carifio, J.; Perla, R.J. Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *J. Soc. Sci.* **2007**, *3*, 106–116. [CrossRef]

33. Carifio, J.; Perla, R. Resolving the 50-year debate around using and misusing Likert scales. *Med. Edu.* **2008**, *42*, 1150–1152. [CrossRef]
34. Kalkbrenner, M.T. Alpha, omega, and h internal consistency reliability estimates: Reviewing these options and when to use them. *Couns. Outcome Res. Eval.* **2021**, *14*, 77–88. [CrossRef]
35. García-Pérez, M.A. Are the steps on Likert scales equidistant? Responses on visual analog scales allow estimating their distances. *Educ. Psychol. Meas.* **2024**, *84*, 91–122. [CrossRef] [PubMed]
36. de Winter, J.F.; Dodou, D. Five-point Likert items: T test versus Mann-Whitney-Wilcoxon. *Pract. Assess. Res. Eval.* **2010**, *15*, 11. [CrossRef]
37. Robbins, N.B.; Heiberger, R.M. Plotting Likert and other rating scales. In Proceedings of the 2011 Joint Statistical Meeting, Miami Beach, FL, USA, 30 July–4 August 2011; American Statistical Association: Alexandria, VA, USA, 2011; Volume 1.
38. Heiberger, R.M.; Robbins, N.B. Design of diverging stacked bar charts for Likert scales and other applications. *J. Stat. Softw.* **2014**, *57*, 1–32. [CrossRef]
39. Bryer, J. Analysis and Visualization of Likert Based Items. Available online: https://jbryer.github.io/likert/ (accessed on 8 November 2024).
40. Heymans, M.W.; Twisk, J.W.R. Handling missing data in clinical research. *J. Clin. Epidemiol.* **2022**, *151*, 185–188. [CrossRef] [PubMed]
41. Resseguier, N.; Giorgi, R.; Paoletti, X. Sensitivity analysis when data are missing not-at-random. *Epidemiology* **2011**, *22*, 282. [CrossRef] [PubMed]