Likert Measurement Scale in Education and Social Sciences: Explored and Explained

Nor Roselidyawaty Mohd Rokeman

Faculty of Technical and Vocational, Sultan Idris Education University, 35900 Tanjong Malim, Perak, Malaysia

email: roselidyawati@moe.edu.my

Abstract

Surveys are fundamental research tools extensively used by researchers and academic experts across various fields. Their significance lies in their ability to give respondents autonomy in evaluating presented items. An essential part of crafting an effective survey involves carefully selecting scaling methods during the instrument construction phase. This study conducts a narrative literature review to analyse existing findings on the use of Likert scales, mapping out how these scales are applied to understand the interplay between scale divisions in social sciences research. Likert scales are popular in diverse fields, including social sciences, psychology, politics, and economics. However, existing methodologies for instrument construction often lack a comprehensive depiction of the most suitable scales for specific studies. This paper aims to address this gap by providing guidance and enhancing understanding for researchers in selecting appropriate and relevant scales, particularly within the context of their research topics. Findings suggest that a more systematic approach to scale selection can improve research efficacy. Recommendations include developing standardized guidelines to aid researchers in choosing optimal scales for their studies, ensuring both methodological rigor and relevance to the research questions.

Keywords: Likert Scale, Social Science, Questionnaire, Survey Design, Response Rate, Rating Scale, Response Scale

INTRODUCTION

Researchers across diverse disciplines, including the education and social sciences, psychology, politics, and economics, utilize various measurement instruments to assess behaviors, characteristics, and opinions to address their research inquiries (Jebb et al., 2021). One widely employed tool in survey-based research designs is the questionnaire. Each respondent's input is expressed through a measurement scale, which captures their behavior or perspectives. The meticulous arrangement of scale divisions as a measuring tool is crucial to minimize errors during data analysis. Once the conceptual framework of a study is established, the choice and application of a specific scale measurement become integral to the research instrument (Taherdoost, 2017).

The Likert scale is extensively used in numerous social science studies (Rokeman & Kob, 2023; Clark & Watson, 2019), particularly in fields such as education (Gupta et al., 2021), business (Schrum et al., 2020), tourism (Bellizzi et al., 2020), and psychology (Necho et al., 2021). It serves to measure respondents' opinions and behaviors. The Likert scale was devised to measure 'attitude' in a scientifically accepted and validated manner (Güler & Ayan, 2020). Attitude is a multidimensional construct involving cognitive, affective, and psychomotor elements. Quantifying these subjective attitudinal components in a validated and reliable way is a challenge (Sijtsma & Ark, 2020). Numerous scholarly debates have emerged regarding the use of the Likert scale, a widely employed tool for measuring such attitudinal constructs. These debates have centered around various aspects, including

the accuracy of the data collected, the ranking order and distance between scale options, the use of multiple scales across different questionnaires, the ability to combine scale options, the appropriateness of performing mathematical operations on Likert scale data, and the quantification of respondents' negation choices (Anjaria, 2022).

The Likert scale is presented as a tool that can help address these challenges by enabling the measurement of attitudinal elements. Researchers face several practical challenges when using Likert scales in data collection and analysis, primarily related to response bias, statistical interpretation, and design inconsistencies. The use of odd versus even response options can influence reliability and validity, with odd scales generally preferred for their effectiveness (Kusmaryono & Wijayanti, 2022). However, issues such as central tendency bias and the misinterpretation of Likert data as interval rather than ordinal can compromise data integrity (Winston, 2021). The effectiveness of the Likert scale is also strongly influenced by the form of the question items in the questionnaire and the data analysis technique employed. Despite its widespread use, a research gap exists concerning the use of scale levels, types of Likert scales, and differences in opinions when adapting the Likert scale in research instruments. This discourse has highlighted two key issues that have emerged as promising areas for further exploration and explanation by educational researchers. These issues warrant in-depth investigation and clear elucidation, as they present valuable opportunities for advancing the understanding of the constructs at hand. This research paper aims to guide and assist researchers in selecting appropriate scale levels for constructing survey scales, facilitating responses that effectively measure research objectives based on the information provided by respondents.

LITERATURE REVIEW

Historical Development of Likert Scales

The Likert Scale was first introduced by Rensis Likert in his 1932 publication "A Technique for Measurement of Attitudes" in the Archives of Psychology journal. This scale is widely utilized in survey-based research, particularly within the social sciences (Bodur et al., 2024; Rokeman & Kob, 2024; Alhassn et al., 2022; Jebb et al., 2021). In his seminal work, Likert (1932) discussed various types of scale measures that can be employed in data collection instruments.

Survey instruments developed by researchers typically present respondents with a series of construct statements, followed by a measurement scale that allows individuals to express their views, attitudes, and responses based on their experiences or opinions regarding these statements. These response options often include several levels of agreement, ranging from "Strongly Agree" to "Strongly Disagree," and may also incorporate mid-range options such as "Agree," "Neutral," and "Disagree." Numerical values or scores are assigned to these response options to quantify and rank the respondents' reactions. By consistently applying this scale across survey items, researchers can effectively measure attitudes and perceptions related to their study. An example of constructs measured on a Likert scale in educational research might include statements related to teaching effectiveness, learning strategies, or student engagement.

Types of Scale Tiers

It is important to understand the various scales of measurement, as a deficiency in this knowledge would obstruct one's ability to apply suitable data analysis techniques. In the context of statistical analysis, numerical values are distributed in accordance with accepted standards, thereby delineating the corresponding measurement scale. Each distinct scale of measurement possesses unique attributes or a collection of attributes that dictate the appropriate statistical techniques to be employed (Allanson et al., 2020). By understanding the level of measurement, one can effectively interpret the data and subsequently ascertain which statistical analyses are appropriate. The classifications of scale tiers pertain to the diverse levels or categories utilized within a measurement scale. The most common types of scale tiers are:

Nominal scale

The nominal scale is considered the lowest level of measurement. This scale assigns a number or label to objects or events, but the numbers do not have any quantitative meaning. They simply serve to categorize or classify the variables into distinct groups without any specific order or ranking. For instance, the variable "gender" can be coded as 1 for male and 2 for female, or it can be represented by the label's "male" and "female." Similarly, marital status can be categorized as "single," "married," or "divorced" without implying any inherent order or numerical relationship between the categories. The nominal scale is the most basic form of measurement and is useful for classifying and organizing data into mutually exclusive groups, but it does not provide information about the magnitude or relative positioning of the variables.

Ordinal scale

The ordinal scale represents a higher level of measurement compared to the nominal scale. This scale indicates the relative position or ranking of items or objects, but the distances between the categories are not necessarily equal. The numbers or labels assigned to the variables have a specific order, but they do not provide information about the magnitude of the differences between them. For example, educational level can be measured on an ordinal scale as "high school," "bachelor's," "master's," and "doctorate," which clearly shows the hierarchical relationship between the levels. Similarly, Likert scales commonly used in surveys, such as "strongly disagree," "disagree," "neutral," "agree," and "strongly agree," represent an ordinal scale that captures the respondents' relative opinions or attitudes, even though the intervals between the response options may not be identical. The ordinal scale is useful for understanding the comparative positioning of variables, but it does not allow for the calculation of precise numerical differences or ratios between the categories.

Interval scale

The interval scale represents a more sophisticated level of measurement compared to the nominal and ordinal scales. This scale has equal intervals between the scale points, which means that the differences between the scale points are meaningful and can be quantified. However, the interval scale lacks a true zero point, which means that ratios and meaningful differences between values cannot be calculated. For example, temperature measured in Celsius or Fahrenheit is an interval scale, as the differences between the temperature readings are equal and can be compared, but the zero point does not represent a complete absence of temperature. Similarly, IQ scores are measured on an interval scale, as the differences between the scores are meaningful, but the zero point does not represent a complete absence of intelligence. The interval scale is useful for understanding the magnitude of differences between variables and for conducting more advanced statistical analyses, but it is limited in its ability to make claims about proportions or ratios between the measured values.

Ratio scale

The ratio scale represents the highest level of measurement. This scale not only has equal intervals between the scale points, like the interval scale, but it also has a true zero point. This means that the ratios between the scale points are meaningful and can be interpreted in a quantitative manner. For example, variables such as height, weight, age, and income are measured on a ratio scale. The ratio scale allows for the most precise and informative quantitative comparisons between variables, as it enables the calculation of ratios, proportions, and meaningful differences. This level of measurement is essential for conducting sophisticated statistical analyses and for drawing robust conclusions in academic research.

Types of Likert Scale

The Likert scale is classified as an ordinal scale (Anjaria, 2022) because it demonstrates ordered levels, yet the distances between scale categories are not uniform. An ordinal scale features ordered levels, but

the precise intervals between these levels remain uncertain (Chaibakhsh & Pourhoseingholi, 2023). For instance, on a Likert scale with five levels, "Strongly Agree" is undeniably higher than "Agree," "Agree" is higher than "Neutral," "Neutral" is higher than "Disagree," and "Disagree" is higher than "Strongly Disagree." However, the distances between "Strongly Agree" and "Agree," from "Agree" to "Neutral," and so forth, are not necessarily equal. Consequently, the data generated by the Likert scale is considered ordinal in nature. Conversely, Raymond et al. (2022) maintains a different perspective, suggesting that the Likert scale is an interval scale. This disagreement between the interval and ordinal views of the Likert scale relates to the use of total scores for each question and individual scores on question information. The summation of scores from all questions is referred to as the test score, which is a focal point in classical test theory (CTT) (Baker, 2001).

In the initial development of the Likert scale, Likert (1932) employed a 5-point scale, with response options ranging from "Strongly Agree" to "Strongly Disagree," including the midpoint "Undecided." The application of score values or scale numbers on Likert-type instruments has shown variability over time, as researchers have explored different methods to enhance the accuracy and validity of data derived from these scales. Likert originally proposed a 5-point scale, but variations have emerged to address challenges in data analysis and interpretation. Commonly used 5-point scales are often treated as interval scales during statistical analyses, despite being ordinal in nature (Mirahmadizadeh et al., 2018). Researchers have utilized both odd- and even-numbered scales, as well as scales extending beyond the original 5-point format, providing respondents with a diverse set of response choices when completing survey instruments.

Table 1. Response Scale Likert (Sources: Simms et al., 2019)

Scale	Description
1	Strongly Disagree
2	Disagree
3	Somewhat Disagree
4	Slightly Disagree
5	Neutral
6	Slightly Agree
7	Somewhat Agree
8	Agree
9	Strongly Agree
10	Completely Agree

Despite the initial use of a symmetrical and balanced five-scale system in early Likert Scale studies, the evolving research landscape has witnessed the expansion of measurement scale points, ranging from as few as two to as many as eleven scale points (Simms et al., 2019). While the variability in Likert-type instruments reflects efforts to improve data accuracy, it also highlights the challenges in achieving consensus on best practices. The ongoing exploration of alternative methods and scale modifications underscores the complexity of effectively utilizing Likert scales in research across various fields. Table 1 provides a summary of the various Likert Scale formats commonly employed across different research domains. This table can be adapted to various contexts, such as customer feedback, employee satisfaction, or research surveys. The descriptions for each scale point can be modified to fit the specific application or research question.

Applications in Different Fields

The Likert scale, a prevalent tool in the social and behavioural sciences, is increasingly applied across various fields, including economics, disaster risk reduction, and organizational resilience. In the education and social sciences, it is used to measure latent variables, though challenges remain in processing ordinal data, which can lead to information distortion. To address this, fuzzy logic-based survey instruments have been developed, allowing for degrees of partial agreement to reduce data loss

(Memmedova & Ertuna, 2024). In economics and other sciences, Likert scales are employed to evaluate characteristics, often viewed as ordered finite sets of categories (Dombi & Jónás, 2021). In disaster risk reduction and organizational resilience, a simplified Likert scale model has been introduced to benchmark operational capacity and resilience. This simplified model is particularly useful in contexts where traditional scales may not be suitable due to cultural or social factors (Pescaroli et al., 2020). Furthermore, in psychometrics, Kernel Smoothing Item Response Theory (KS-IRT) offers a nonparametric approach to evaluate Likert items, providing graphical insights into item performance across different trait levels (Baghaei & Effatpanah, 2024). Despite its widespread use, challenges remain in designing and analyzing Likert scale data, particularly in determining the appropriate statistical tests (Mirahmadizadeh et al., 2018). The Likert scale's adaptability and simplicity make it a valuable tool across diverse disciplines, though ongoing innovations and methodologies continue to enhance its application and accuracy.

METHODOLOGY

This literature review employs a comprehensive analysis of previous studies by examining a diverse range of relevant documents, including journal articles, seminar proceedings, magazines, newspapers, books, and online sources (Galvan, 2017) using a narrative reviews steps by Demiris et al. (2019). Given the explanatory nature of this research, the study relies on secondary data collection methods. Secondary data, defined as information "previously collected and tabulated by other sources" (Bhattacharjee, 2012), is utilized to gain a deeper understanding of the research subject and to examine the work of previous researchers in the same domain. The study draws upon textbooks, internet-based resources, and electronic databases to establish the necessary scientific background and to access relevant journals and articles.

This paper synthesizes research findings and related literature, providing potential explanations for these findings and briefly discussing implications for both research and practice. The presentation of findings follows a chronological order, employing a narrative review approach. This narrative review offers a broad overview of the general topic studied, allowing for a comprehensive exploration of themes and trends over time (Sukhera, 2022). By utilizing this methodology, the study aims to provide a thorough and well-structured analysis of the existing literature, contributing to a deeper understanding of the research topic and its implications in both theoretical and practical contexts.

Search strategy

The search strategy approach emphasizes participatory action research (PAR), which fosters collaboration between researchers and participants, enhancing the relevance and applicability of findings in real-world settings. The aim of this review was to explore the following three issues are addressed: 1) How many Likert-type items are needed for a Likert attitude scale? 2) What and how many response categories should be presented? Specifically in education and social sciences research settings. Demiris et al. (2019) provides a comprehensive guide on conducting narrative reviews. It outlines a step-by-step approach, including:

Step 1: Literature Search

A comprehensive literature search was conducted to extract relevant research from published sources. Multiple electronic databases were utilized to ensure a wide retrieval of pertinent studies. The databases searched included:

- Scopus
- Google
- Google Scholar

Google Scholar was specifically included as it is a subset of the more extensive Google search index and comprises full-text journal articles, technical reports, preprints, theses, books, and other scholarly documents, including selected web pages deemed scholarly (Vina, 2006). The inclusion of both

traditional academic databases and broader search engines like Google ensured a thorough exploration of available literature.

Step 2: Identification of Keywords

To guide the search process and maximize the retrieval of relevant studies, a set of keywords and phrases were identified and employed. The primary keywords used were "Likert scale types," "social sciences," "education," and "categories." These were used both individually and in combination with additional terms and synonyms to expand the search scope. Combinations and variations included "types of Likert scales," "response categories," "social sciences Likert scale," and "application of Likert scales." This comprehensive approach to keyword selection ensured a wide-ranging exploration of the literature, capturing various aspects of Likert scale usage across different disciplines, particularly within social sciences and education. Synonyms and related phrases were incorporated as appropriate to capture a comprehensive array of literature on the topic. This strategic use of keywords was essential to identify studies that specifically address the number and structure of Likert items and response categories within the contexts of social sciences and education.

Step 3: Screening and Selection of Articles

The search was refined using specific inclusion criteria to select studies most relevant to the research questions. The criteria were as follows:

- **Time Frame:** Publications from **2000 to 2024** were considered. This period was chosen to focus on recent developments and trends in Likert scale research, recognizing the rapidly changing nature of social science research methodologies in recent years. The year 2000 was selected as the starting point due to sociocultural shifts influencing research practices in social sciences and education.
- **Subject Matter:** The studies had to involve the use of the Likert scale within social sciences or educational research contexts.
- **Publication Type:** Publications included peer-reviewed journal articles, books, theses, and conference proceedings.

Following the search, all retrieved articles were compiled, and duplicate entries were removed. The abstracts of the remaining articles were reviewed to assess their relevance to the research questions. Articles that met the inclusion criteria were selected for further analysis. According to Ahn and Kang (2018), inclusion criteria are essential for defining the general characteristics of research subjects from a target population to be studied.

Step 4: Data Extraction and Synthesis

From the selected articles, key information was extracted and synthesized to address the research objectives. This process involving highlighting the key information from the selected articles was meticulously extracted and synthesized. This process began with summarizing the main concepts, where central themes and findings related to Likert scale construction and use were highlighted. Methodological analysis followed, examining the research methodologies employed in the studies, with a particular focus on how Likert scales were utilized, including the number of items and response categories. The analysis also identified prevalent trends in the number of Likert-type items and response categories used in recent research, as well as gaps in the literature that the current study could address. Finally, the findings were integrated to provide a cohesive narrative that informs the current study's approach to using Likert scales in social sciences and educational research. This comprehensive synthesis ensures that the current study is grounded in existing literature while addressing identified gaps and trends.

FINDINGS AND DISCUSSIONS

Optimal Number of Points on Likert Scales in Social Sciences and Education Research: A Comparative Analysis of Five-, Seven-, and Ten-Point Scales

The appropriate selection of rating scales is crucial because it can influence the scale categories and subsequent responses from participants (Preston & Colman, 2000). In this context, the seven-point Likert scale has been shown to correlate significantly better than the five-point scale (Finstad, 2010). Finstad (2010) argues that a seven-point scale is more likely to capture genuine responses from survey respondents. This observation is further supported by Marco et al. (2024), who found that the seven-point scale is generally preferred by respondents due to its more precise and user-friendly nature. Thus, the choice between a five-point and a seven-point scale should consider the specific context and the psychometric performance of each scale in capturing genuine responses. However, some researchers suggest that the five-point scale may be less confusing for respondents and might even enhance response rates (Russo et al., 2021). Regarding the use of five, seven, and ten-point scales, Preston and Colman (2000) suggest that five and seven-point scales are the most user-friendly, although ten- and eleven-point scales can be suitable for tapping into the feelings or opinions of study respondents.

Taherdoost (2019) points out that a scale that is too short cannot adequately capture differences in opinions, while a Likert scale with too many points, such as a 12-point scale (McRae, 1970), can dilute the information conveyed by the instrument. Conversely, Awang et al. (2016) propose that a ten-point Likert scale can yield accurate measurement results for Structural Equation Modelling (SEM) analysis in parametric tests, as both measurement models and structures can be assessed more effectively. Additionally, Sangthong (2020) suggests that scales with 5, 7, and 10 points, used with sample sizes of 15, 30, 50, and 100, can effectively control Type 1 errors under F-test and K-W test conditions, assuming a normal population and positive skewness values.

Furthermore, researchers must consider the potential for social bias when assigning scale values in survey statements, as there may be differences in bias between odd-numbered Likert scales (e.g., five points) and even-numbered Likert scales (e.g., four points) (Kusmaryono et al., 2022). However, Preston and Colman (2000) note that using scales with fewer than five points has drawbacks, such as being less reliable, valid, discriminative, and stable. Finstad (2010) also critiques the five-point scale for being less sensitive in obtaining user responses, while viewing the seven-point scale as more suitable for obtaining accurate assessments of research information.

In exploratory research, optimal scale points for measuring attitudes often favour shorter scales, typically five to seven points, to facilitate respondent engagement and reduce cognitive load. This approach captures a range of attitudes without overwhelming participants, which is crucial in initial investigations where constructs are not fully defined (Goggin & Stoker, 2014). In contrast, confirmatory research tends to utilize longer scales, often seven to eleven points, to enhance measurement precision and validate hypothesized relationships among constructs (Yoon, 2024). This distinction arises because confirmatory studies require more nuanced data to test specific theories, benefiting from the additional granularity that longer scales provide. While shorter scales are advantageous in exploratory contexts, they may lack the depth required for confirmatory analysis, highlighting the trade-off between ease of response and measurement precision.

Controversies in Response Option Selection

The Likert scale, a staple in social science research and survey design since its introduction by Rensis Likert in 1932, serves as a fundamental tool for measuring attitudes, perceptions, and behaviours. By asking respondents to indicate their level of agreement or disagreement with a series of statements, researchers can quantify subjective phenomena and analyse trends across diverse populations. However, despite its widespread adoption, the construction of Likert scales—particularly the selection of response options—has been a subject of ongoing controversy and debate among scholars and practitioners. These controversies surrounding response option selection are among the key issues debated by researchers in the field of survey methodology:

Ordinal Scale or Interval Scale Points

Ordinal scales rank data without consistent intervals, while interval scales provide equal distances between values, allowing for more robust statistical analysis. This distinction is crucial for ensuring reliability and validity in measurement, as unreliable tools can misrepresent true values. The interpretation of scale types can also lead to confusion in research practices, emphasizing the need for a well-developed theoretical context to establish measurement validity (Feuerstahler, 2023). Additionally, improper treatment of ordinal data as continuous in statistical modelling, such as structural equation modelling, can yield biased results (Robitzsch, 2022). Choice between odd and even scale points might depend on whether the goal is to allow for a neutral midpoint or to guide respondents toward a particular side, with an even scale potentially serving the latter purpose better. Thus, these measurement debates have profound implications for both theoretical development and empirical research outcomes.

Impact of Response Categories on Psychometric Quality

The use of scales on item statements in survey instruments is often scrutinized for their psychometric quality and potential systematic measurement errors (Menold & Bogner, 2016). The validity and reliability of these scales are determined based on anticipated responses. Contrary to the notion that the number of response categories impacts respondents' answers, commonly used measurement scales typically consist of six or seven points (Simms et al., 2019). Some researchers endorse the use of a seven-point scale due to its ability to provide high reliability (Sirganci & Uyumaz, 2022). Controversies in Likert scale response options primarily stem from issues related to item format, response bias, and psychometric properties.

Reversed Item Formats and Misresponse Issues

The use of reversed items is particularly contentious, as it can lead to misresponse, where respondents select answers on the same side of the neutral point for both reversed and non-reversed items, complicating data interpretation (Winston, 2021). Additionally, the number of response options can affect the reliability and validity of scales, with findings suggesting that fewer than six options may diminish psychometric precision (Simms et al., 2019). Mixed item formats (positive and negative) can obscure significant differences in responses, leading to misleading conclusions (García-Fernández et al., 2022). Moreover, mixed item formats can complicate the factor structure during data analysis. For instance, positively and negatively worded items intended to measure the same construct might load onto separate factors in exploratory factor analyses, not because they represent different underlying constructs but due to the wording direction. This misrepresentation can lead researchers to misleading conclusions about the constructs being measured. It may suggest multidimensionality where there is none, thereby affecting the validity of the scale.

Scale Length and Psychometric Precision

The controversies surrounding Likert scale design underscore the inherent complexities in achieving both reliability and validity in survey research. While some researchers advocate for flexibility in response options to enhance data richness, this approach necessitates meticulous scale design to mitigate potential issues. The application of Cronbach's alpha for internal consistency reveals that seven- to ten-point scales yield higher reliability, whereas three-point scales result in the lowest Cronbach alpha readings (Sirganci & Uyumaz, 2022; Hartley, 2013). This suggests that increasing the number of scale points can enhance reliability. Additionally, higher scale options have been associated with increased validity (Wakita et al., 2012), further supporting the use of more granular scales.

However, the preference for higher scale points must be balanced against practical considerations, such as respondent burden and the potential for increased response bias. While higher scales may offer greater sensitivity and the ability to capture nuanced attitudes, they can also complicate the response process and lead to central tendency bias. Therefore, the choice of scale points should be informed by the specific research context and objectives, ensuring that the benefits of increased

reliability and validity are not offset by practical drawbacks. This critical analysis highlights the need for a nuanced approach to Likert scale design, one that carefully considers the trade-offs between scale granularity and respondent engagement.

Respondent Preferences for Scale Point Numbers

Respondent preferences for the number of scale points in Likert scales vary significantly based on factors such as ease of use, reliability, and the ability to express nuanced attitudes. Research indicates that while five-point scales are the most commonly used, scales with more points can enhance sensitivity and reliability. Preston and Colman (2000) found that respondents favour 5, 7, and 10-point scales for their ease of use. However, two-, three-, and four-point scales facilitate quick choices, providing sufficient points to express agreement and opinion. Nonetheless, the ten-point scale, followed by the seven-point and nine-point scales, are preferred by most respondents. Most researchers prefer a 5-point Likert scale, while a 7-point scale is deemed most effective for reliability and validity, with even-numbered scales used to direct responses (Kusmaryono et al., 2022). Thus, Likert scales with odd responses over five points are effective. However, response bias and central tendency bias can affect the validity and reliability of these scales.

Odd versus Even Numbered Scale Points

The use of odd or even numbers as scale points in Likert scales is a consideration. While James (2019) suggests respondents prefer odd-numbered scale points, there is no specific recommendation as they do not significantly impact psychometric measurements (Menold & Bogner, 2016). Researchers may choose to arrange measurement scales in ascending or descending order (Menold & Bogner, 2016). Choice between odd and even scale points might depend on whether the goal is to allow for a neutral midpoint or to guide respondents toward a particular side, with an even scale potentially serving the latter purpose better. The researcher recommended to use a seven-point scale with an odd number of responses. Nevertheless, should the investigator seek to guide the participant towards a particular perspective, then employing a six-point scale characterized by an even number of options might prove to be more appropriate.

CONCLUSION

The Likert scale is a widely used tool for collecting research data, offering accurate results when appropriately applied. However, the optimal number of Likert scale points depends on the specific study design and the desired responses from respondents. Respondents generally find it easier to select scale points when survey statements are clear and not confusing. Thus, a meticulous arrangement of the Likert scale is necessary to avoid errors in data analysis.

For instance, Awang et al. (2016) propose that a ten-point Likert scale can yield accurate measurement results for Structural Equation Modelling (SEM) analysis in parametric tests, as both measurement models and structures can be effectively assessed. This scale is expected to be more successful in determining construct validity. Additionally, Sangthong (2020) suggests using scales like 5, 7, and 10 points with appropriate sample sizes to help control Type 1 errors, considering F-test and K-W test conditions based on normal population assumptions and positive skewness values. For studies requiring high reliability and precision, a 7-point scale is recommended. This scale offers a balance between simplicity for respondents and detailed data capture. In exploratory studies with large sample sizes, employing a 10-point scale may provide more nuanced insights into participant attitudes and opinions, particularly in complex analyses such as SEM.

Furthermore, researchers must be mindful of potential social bias when assigning scale values in survey statements, as there may be differences in bias between odd Likert scales (like 5 points) and even Likert scales (like 4 points). However, scales with fewer than 5 points have drawbacks, as they are less suitable for reliability, validity, discrimination, and stability. In summary, the Likert scale can yield accurate research data when its design is carefully considered to match study objectives and respondent preferences. Researchers should strive to create clear and unambiguous survey statements while

addressing potential sources of bias in scale selection to ensure the validity and reliability of their research findings.

RECOMMENDATIONS FOR FUTURE RESEARCH

The use of the Likert scale has become increasingly prevalent in social science research over the past decade. Researchers often rely on this scale as a crucial tool for analysing data and addressing issues raised in their studies. However, a common oversight is the lack of attention given to the necessary considerations for its proper implementation, which can potentially lead to data bias. To address this concern, the findings of the present research propose alternative solutions to guide the effective use of the Likert scale in social science research. These recommendations aim to enhance the reliability and validity of the data collected using this scale. For instance, online survey tools facilitate the distribution of Likert scale surveys to a broader audience, thereby improving data collection efficiency. Additionally, these digital platforms offer customizable options for Likert scale design, such as varying response formats and visual elements, which improve user engagement. These advancements can significantly influence how researchers design and implement Likert scales in their studies.

Furthermore, it is essential to conduct additional studies on the use of the Likert scale across various aspects analysed in specific fields. We provide several recommendations for compiling and analysing survey questionnaires to minimize bias and potential problems:

1. When designing a Likert scale questionnaire, use a balanced key by ensuring the number of positive questions equals the number of negative statements.

2. Include more question items in the questionnaire, using an odd-numbered response scale (5, 7, 9, or 11 points). Conduct data analysis appropriately based on the type of data (ordinal or interval) and avoid using the mean for ordinal data. Analyse all Likert items provided to participants. Once you have settled on an analysis plan, apply it comprehensively to all collected Likert data and report results for each question. Reporting results for only a few questions can weaken credibility and be misleading.

Implementing these recommendations would provide a more comprehensive understanding of the scale's applicability and potential limitations, ultimately contributing to the improvement of research methodologies in social science disciplines. By acknowledging the prerequisites for using the Likert scale and exploring alternative approaches, researchers can optimize the implementation of this tool to ensure the integrity and robustness of their research findings. Continued investigation into the Likert scale's applications across diverse education and social science domains will further enhance methodological rigor and improve the overall quality of research in this field.

REFERENCES

Ahn, E., & Kang, H. (2018). Introduction to systematic review and meta-analysis. *Korean Journal of Anaesthesiology*, *71*(2), 103-112. Doi: https://doi.org/10.4097/kjae.2018.71.2.103

Allanson, Patricia E., and Notar, Charles E. (2020). Statistics as Measurement: 4 Scales/Levels of Measurement. *In: Education Quarterly Reviews*, *3*(3), 375-385. DOI: https://doi.org/10.31014/aior.1993.03.03.146

Alhassn, I., Asiamah, N., Opuni, F. F., & Alhassan, A. (2022). The Likert scale: exploring the unknowns and their potential to mislead the world. *UDS International Journal of Development*, *9*(2), 867-880.
Doi: https://doi.org/10.47740/586.UDSIJD6i

Anjaria, K. (2022). Knowledge derivation from Likert scale using Z-numbers. *Information Sciences*, *590*, 234-252. Doi: https://doi.org/10.1016/j.ins.2022.01.024

Awang, Z., Afthanorhan, A., & Mamat, M. (2016). The Likert scale analysis using parametric based Structural Equation Modeling (SEM). *Computational Methods in Social Sciences, 4*(1), 13-21.

Baghaei, P., & Effatpanah, F. (2024). Nonparametric Kernel Smoothing Item Response Theory Analysis of Likert Items. *Psych*, *6*(1), 236-259. Doi: https://doi.org/10.3390/psych6010015

Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC.

Bhattacharjee, S., Rajaraman, P., Jacobs, K. B., Wheeler, W. A., Melin, B. S., Hartge, P., ... & Chatterjee, N. (2012). A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *The American Journal of Human Genetics*, *90*(5), 821-835.

Bellizzi, M. G., Eboli, L., & Mazzulla, G. (2020). Air Transport Service Quality Factors: A Systematic Literature Review. *Transportation Research Procedia, 45*, 218-225. Doi: https://doi.org/10.1016/j.trpro.2020.03.010

Bodur, S., Topal, S., Gürkan, H., & Edalatpanah, S. A. (2024). A Novel Neutrosophic Likert Scale Analysis of Perceptions of Organizational Distributive Justice via a Score Function: A Complete Statistical Study and Symmetry Evidence Using Real-Life Survey Data. *Symmetry*, *16*(5), 598. Doi: https://doi.org/10.3390/sym16050598

Chaibakhsh, S., & Pourhoseingholi, A. (2023). A Modification on Intra Class Correlation Estimation for Ordinal Scale Variable Using Latent Variable Model. *Journal of Biostatistics and Epidemiology*. Doi: https://doi.org/10.18502/jbe.v9i1.13972

Clark, L. A., & Watson, D. (2019). Constructing Validity: New Developments in Creating Objective Measuring Instruments. *Psychol. Assess. 31*, 1412. Doi: https://doi.10.1037/pas0000626

Demiris, G., Parker Oliver, D., & Washington, K. (2019). *Narrative review: A guide for researchers*. Publisher.

Dombi, J., Jónás, T., Dombi, J., & Jónás, T. (2021). Likert Scale-Based Evaluations with Flexible Fuzzy Numbers. *Advances in the Theory of Probabilistic and Fuzzy Data Scientific Methods with Applications*, 167-187. Doi: https://doi.org/10.1007/978-3-030-51949-0_5

Feuerstahler, L. (2023). Scale type revisited: Some misconceptions, misinterpretations, and recommendations. *Psych*, *5*(2), 234-248. Doi: https://doi.org/10.3390/psych5020018

Finstad, K. (2010). Response Interpolation and Scale Sensitivity: Evidence against 5-point scales. *Journal of Usability Studies, 5*(3), 104-110.

Galvan, J. L. (2017). *Writing Literature Review a Guide for Students of the Social and Behavioural Science*. Oxon: Routledge.

García-Fernández, J., Postigo, Á., Cuesta, M., González-Nuevo, C., Menéndez-Aller, Á., & García-Cueto, E. (2022). To be direct or not: Reversing Likert response format items. *The Spanish Journal of Psychology*, *25*, e24. Doi: https://10.1017/SJP.2022.20

Goggin, S., & Stoker, L. (2014). Optimal scale length and single-item attitude measures: evidence from simulations and a two-wave experiment. In *APSA 2014 Annual Meeting Paper*.

Güler, G., & Ayan, C. (2020). Review of Attitude Scales Developed in Turkey Between 2002-2018 Regarding the Scale Development Process. *Journal of Faculty of Educational Sciences*, *53*(3). Doi: DOI: https://10.30964/auebfd.658488

Gupta, Y., Khan, F. M., & Agarwal, S. (2021). Exploring Factors Influencing Mobile Learning in Higher Education-A Systematic Review. *International Journal of Interactive Mobile Technologies, 15*(12). Doi: https://doi.org/10.3991/ijim.v15i12.22503

Hartley, J. (2013). Some Thoughts on Likert-type scales. *Int J Clin Health Psychology, 13*, 83-86. Doi: https://doi.org/10.1016/S1697-2600(14)70040-7

James, R. L. (2019). Measuring user experience with 3, 5, 7, or 11 points: Does it matter? Human Factors: The Journal of the Human Factors and Ergonomics Society, 63(6), 999–1011. https://doi.org/10.1177/0018720819881312

Jebb, A. T., Ng, V., & Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Frontiers in Psychology, 12*, 637547. https://doi.org/10.3389/fpsyg.2021.637547

Johns, R. (2010). Likert Items and Scales. *SQB*.

Kusmaryono, I., Wijayanti, D., & Maharani, H. R. (2022). Number of Response Options, Reliability, Validity, and Potential Bias in the Use of the Likert Scale Education and Social Science Research: A Literature Review. *International Journal of Educational Methodology*, *8*(4), 625-637. Doi: https://doi.org/10.12973/ijem.8.4.625

Likert, R. (1932). A Technique for The Measurement of Attitudes. (R. Woodworth, Ed.) *Archives of Psychology, 140*, 5-55.

Marco, A., Núñez, R., Ramírez, K., Alejandra, G.-R., Irma, G., & Esparza, G. (2024). Measurement invariance in five-point and seven-point Likert scale of the SWLS in five Ibero-American countries. *Psicumex*, *14*(1). https://doi.org/10.36793/psicumex.v14i1.605

McRae, A. W. (1970). Channel Capacity in Absolute Judgment Tasks: An Artifact of Information Bias? *Psychological Bulletin, 73*, 112-121. Doi: https://psycnet.apa.org/doi/10.1037/h0028513

Memmedova, K., & Ertuna, B. (2024). Development of a fuzzy Likert scales to measure variables in social sciences. *Information Sciences*, *654*, 119792. Doi: https://doi.org/10.1016/j.ins.2023.119792

Menold, N., & Bogner, K. (2016). *Design of Rating Scales in Questionnaires*. GESIS survey guidelines, 4. Doi: https://10.15465/gesis-sg_en_015

Mirahmadizadeh, A., Delam, H., Seif, M., & Bahrami, R. (2018). Designing, constructing, and analyzing Likert scale data. *Journal of Education and Community Health*, *5*(3), 63-72. Doi: https://doi.org/10.21859/jech.5.3.63

Necho, M., Tsehay, M., Birkie, M., Biset, G., & Tadesse, E. (2021). Prevalence of Anxiety, Depression, And Psychological Distress Among the General Population During the COVID-19 Pandemic: A Systematic Review and Meta-Analysis. *International Journal of Social Psychiatry, 67*(7), 892-906. Doi: https://doi.org/10.1177/00207640211003121

Pescaroli, G., Velazquez, O., Alcántara-Ayala, I., Galasso, C., Kostkova, P., & Alexander, D. (2020). A Likert scale-based model for benchmarking operational capacity, organizational resilience, and disaster risk reduction. *International Journal of Disaster Risk Science*, *11*, 404-409. Doi: https://doi.org/10.1007/s13753-020-00276-9

Preston, C. C., & Colman, A. M. (2000). Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, And Respondent Preferences. *Acta Psychological, 104*(1), 1-15. Doi: https://doi.org/10.1016/S0001-6918(99)00050-5

Raymond, Doe., Bryan, A., Landrum., Kaelyn, M., Lewis., Matthew, E., Glenn., Jacob, D., Smith. (2022). Likert Items: Should(n't) We Really Care? *International journal of social science studies, 11*(1):32-32. https://doi.org/10.11114/ijsss.v11i1.5747

Robitzsch, A. (2022). On the bias in structural equation modeling when treating discrete variables as ordinal instead of continuous. *PsyArXiv, 1-16.* https://doi.org/10.3390/axioms11040162

Rokeman, N. R. M., & Kob, C. G. C. (2023). Tahap Kepuasan Kerja Pensyarah Vokasional Terhadap Pencapaian Prestasi di Kolej Vokasional Malaysia: Satu Kajian Rintis. *Jurnal ILMI*, *13*(1), 124-139.

Rokeman, N. R. M., & Kob, C. G. C. (2024). Evaluation of the Content Validity Index of a Job Satisfaction Instrument on Performance among Tourism Lecturers in a Malaysian Vocational College. *International Journal of Academic Research in Business and Social Sciences*, *14*(2), 1669–1683. Doi: http://dx.doi.org/10.6007/IJARBSS/v14-i2/20965

Russo, G. M., Tomei, P. A., Serra, B., & Mello, S. (2021). Differences in the use of 5-or 7-point likert scale: an application in food safety culture. *Organizational Cultures*, *21*(2), 1. Doi: https://10.18848/2327-8013/CGP/v21i02/1-17

Sangthong, M. (2020). The Effect of The Likert Point Scale and Sample Size on The Efficiency of Parametric and Nonparametric Tests. *Thailand Statistician, 18*(1), 55-64.

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the Number of Response Options Matter? Psychometric Perspectives Using Personality Questionnaire Data. *Psychological Assessment*, 1-9. Doi: https://doi.org/10.1037/pas0000648

Sijtsma, K., & van der Ark, L. A. (2020). *Measurement models for psychological attributes*. CRC Press.

Sirganci, G., & Uyumaz, G. (2022). Determining the factors affecting the psychological distance between categories in the rating scale. *International Journal of Contemporary Educational Research*, *8*(3), 178-190. Doi: https://doi.org/10.33200/ijcer.858599

Schrum, M. L., Johnson, M., Ghuy, M., & Gombolay, M. C. (2020). Four years in review: Statistical practices of Likert scales in human-robot interaction studies. In Companion of the 2020 ACM/IEEE *International Conference on Human-Robot Interaction* (pp. 43-52). Doi: https://doi.org/10.1145/3371382.3380739

Sukhera, J. (2022). Narrative reviews in medical education: key steps for researchers. *Journal of Graduate Medical Education*, *14*(4), 418-419. Doi: https://doi.org/10.4300/JGME-D-22-00481.1

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

Vine, R. (2006). Google Scholar. *Journal of the Medical Library Association, 94*(1), 97–99.

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. Doi: https://doi.org/10.1177/0013164411431162

Winston, B. E. (2021). A critique of Likert-response items in social science research: A research note. In *Handbook of Research on Advancements in Organizational Data Collection and Measurements: Strategies for Addressing Attitudes, Beliefs, and Behaviour,* 43-50. IGI Global. Doi: https://10.4018/978-1-7998-7665-6.ch003

Yoon, G. (2024). No One Optimal Way to Measure People's Attitudes? Preferred Length of Scales in Advertising Research. *Journal of Current Issues & Research in Advertising*, *45*(1), 43-70. Doi: https://doi.org/10.1080/10641734.2023.2246049