


FEATURE ARTICLE OPEN ACCESS

A Guide to Key Decision Criteria for Likert-Scale Use in Survey Research

Khan Md Raziuddin Taufique^{1,2} | Md. Mahiuddin Sabbir³ | Fazlul K. Rabbanee⁴

¹Oxford Brookes Business School, Oxford Brookes University, Oxford, UK | ²Faculty of Business Administration, American International University-Bangladesh, Dhaka, Bangladesh | ³Department of Marketing, Faculty of Business Studies, University of Barishal, Barishal, Bangladesh | ⁴School of Management and Marketing, Curtin University, Bentley, Perth, Australia

Correspondence: Fazlul K. Rabbanee (F.Rabbanee@curtin.edu.au)

Received: 28 December 2024 | **Revised:** 12 February 2026 | **Accepted:** 16 February 2026

Keywords: marketing research | questionnaire items | rating scale | response bias | survey method

ABSTRACT

Although widely used in survey research, the application of the Likert scale often lacks rigorous justification in relation to key methodological decisions. Furthermore, inconsistencies in terminology persist, for example, the common reference made to the “5-point Likert scale,” even though debate is ongoing about the optimal number of response categories. Using a narrative review approach, this paper provides researchers with a concise, evidence-based guide for the appropriate use of Likert-type scales in survey research. Drawing on existing literature on itemized rating scales, particularly the Likert scale, the paper identifies the frequently overlooked critical decision criteria for anchoring the Likert scale. It clarifies essential considerations such as the number of scale points, inclusion of midpoints, labeling of response categories, and response format presentation. As a practical contribution, the paper offers evidence-based guidelines to help researchers make informed decisions when employing Likert-type rating scales in survey research.

1 | Introduction

Researchers frequently use rating scales to measure latent variables such as attitudes, perception, personality, and willingness to perform a specific behavior (see e.g., Lim et al. 2025; Paramitha et al. 2025; Presbitero et al. 2025; Soon and Lim 2025; Taufique et al. 2024; Yadav and Kar 2024). Several alternative scale formats are available, including the continuous scale, Likert scale, semantic differential scale, Stapel scale, and Thurstone differential scale. However, the Likert scale, introduced by Rensis Likert (Likert 1932), is one of the most widely adopted instruments in social science (Alabi and Jelili 2023; Chyung et al. 2017) and market research (Heo et al. 2022). The popularity of the Likert scale in social science research is attributed to a number of factors. First, the Likert scale is relatively straightforward to develop and adapt, while being easily comprehensible for respondents,

although its reliability is not without question (Heo et al. 2022), with some of these factors discussed later in this paper. Second, data collected through the Likert scale can be effectively analyzed using statistical tools. Finally, measurements based on the Likert scale have demonstrated commendable validity (Li 2013). The extensive use of the Likert scale is also due to the popularity of self-reported measures in broad social science research (Maulana et al. 2019; Miyatake and Higuchi 2017) and behavioral studies (Sabbir et al. 2023; Sabbir et al. 2021; Sarah et al. 2021).

When using a Likert scale, respondents are asked to indicate the extent to which they agree or disagree with a declarative statement. In a typical 5-point Likert-type scale, for example, response categories are labeled as 1 = “strongly disagree,” 2 = “disagree,” 3 = “neither disagree nor agree,” 4 = “agree,” and 5 = “strongly agree.” However, scale labeling can be altered depending on what

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDeriv](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2026 The Author(s). *Global Business and Organizational Excellence* published by Wiley Periodicals LLC.

is being measured. For example, “never”—“always” labels are used for assessing frequency.

The term “Likert scale” is occasionally employed interchangeably with the term “5-point Likert scale”, implying a default assumption of five response categories (AL-Abrow et al. 2021; Dolnicar 2021; Gomes et al. 2023; Haque et al. 2024; Makanyeza et al. 2023; Oliveira et al. 2023). However, certain studies (Hossain 2019; Lim et al. 2023) adopted a 7-point scale as an alternative. Weijters et al. (2010) presented an overview of the formats commonly employed in marketing research, indicating that the predominant scale format in marketing research is the 7-point scale (55%), followed by the 5-point scale (30%). Both the 5-point and 7-point Likert scales typically feature labeling at extreme points, with the midpoint represented by an odd number. Although these two formats emerge as predominant, no conclusive standardization of the response scale formats is found across studies (Weijters et al. 2010). Evidence in the extant literature has even been found of the use of a 6-point Likert scale which does not have a midpoint (Taufique 2022; Taufique et al. 2017). Clearly, consensus is lacking on the ideal number of response categories, with formats including those featuring five, six, or seven response categories, either fully labeled (i.e., all response categories possess a verbal description) or solely labeled at the extremes or endpoints (e.g., labeling the initial category as “strongly disagree” and the final category as “strongly agree”) (Bearden and Netemeyer 1999; Bruner et al. 2001).

Addressing these issues is crucial as they are highly relevant to marketing and consumer behavior research, which relies heavily on Likert scale-based surveys, particularly those that use cross-sectional designs. Most of these studies depend on self-reported responses, making the choice of an appropriate scale format critical to minimize response bias. In an attempt to reduce these types of bias, researchers sometimes used different scale formats within a questionnaire (Podsakoff et al. 2024) without fully considering the appropriate criteria or contextual factors. For example, researchers may reduce the number of scale points (e.g., yes, no, do not know, etc.) to mitigate response fatigue (Tuan et al. 2005), neglecting other psychometric properties such as reliability and validity. With the dramatic rise in internet usage, online surveys have become increasingly common. Yet responses can vary depending on whether questions are presented vertically or horizontally—an issue marketing and consumer behavior researchers have often failed to consider (Weijters et al. 2021). To date, a handful of studies (see Appendix A1) have examined these aspects of Likert-scale design.

While the current findings (see Appendix A1) offer valuable insights, they focus on a single aspect (e.g., the number of response points) of Likert-scale-related issues and do not provide comprehensive, quick guidelines for broad business and management researchers, especially in the marketing and consumer behavior domains. This study, therefore, aims to fill this gap by synthesizing existing research and offering clear, context-driven guidance on critical scale-design elements to enhance measurement reliability and validity. Specifically, this paper aims to highlight four critical decision criteria for the appropriate use of Likert-type scales in marketing and consumer behavior research: (a) the number of response categories (i.e., scale points); (b) the

use of midpoints; (c) the labeling of response categories; and (d) the format of response categories (horizontal vs. vertical).

2 | Method

This paper adopts a narrative review approach (Baethge et al. 2019) to identify and synthesize the most relevant articles on the topic of interest (Demiris et al. 2019). Unlike systematic reviews, narrative reviews do not follow a rigid, predefined protocol; instead, they provide an evidence-based synthesis based on reviewers’ judgement. This approach is particularly suitable when the existing body of literature is broad and diverse (Nørfelt et al. 2025; Sukhera 2022), which is the case for the subject matter addressed in this paper. Narrative reviews also allow authors to incorporate interpretation and critical insight while offering a comprehensive overview of the key topic (Sukhera 2022). Consistent with this purpose, we reviewed and synthesized the most relevant studies to identify and clarify the decision criteria for the appropriate use of Likert-type scales in survey research (Sukhera 2022; van Dinther et al. 2011).

We developed the search string based on existing literature on Likert-scale measurement, using the following key words: “Likert scale” OR “Likert-type scale” OR “rating scale” OR “itemized rating scale” OR “measurement scale” OR “response style.” Using the Scopus database, we searched for documents containing these keywords in the article titles. Scopus was selected because of its extensive journal coverage (more than 29,100 titles, <https://www.elsevier.com/en-in/products/scopus/content#0-content-coverage>) and its frequent use in similar review studies (e.g., ShabbirHusain et al. 2025; Tortajada et al. 2024). To ensure precision and relevance, the search was purposely restricted to article titles, as our objective was to identify methodological papers on Likert scales rather than studies that merely used these scales as a tool. To maintain a focus on business and management research, we further limited the search to English-language articles published in journals listed in the Australian Business Deans Council (ABDC) 2022 journal ranking, which is a widely recognized and peer-reviewed classification of academic journals in business, management and related fields. This inclusion criterion is consistent with established practices in comparable reviews (Dhaliwal et al. 2025; Gupta et al. 2023). The search resulted in 395 retrieved records.

We first screened the retrieved records by reading their abstracts and retained only those methodological articles that specifically addressed Likert or Likert-type scales and response styles. This resulted in a final selection of 27 articles from 23 journals (see Appendix A1), all of which contributed to understanding response bias and the decision criteria for using Likert and other itemized rating scales in business and management research. Following the procedure used in prior work (e.g., Gupta et al. 2023), the first author conducted an initial, in-depth review of the selected papers and extracted the key findings. All authors then engaged in a collective discussion to confirm the appropriateness and relevance of these findings for the current study. The results of this review are presented in the following section, and the overall methodological process is summarized in Figure 1.

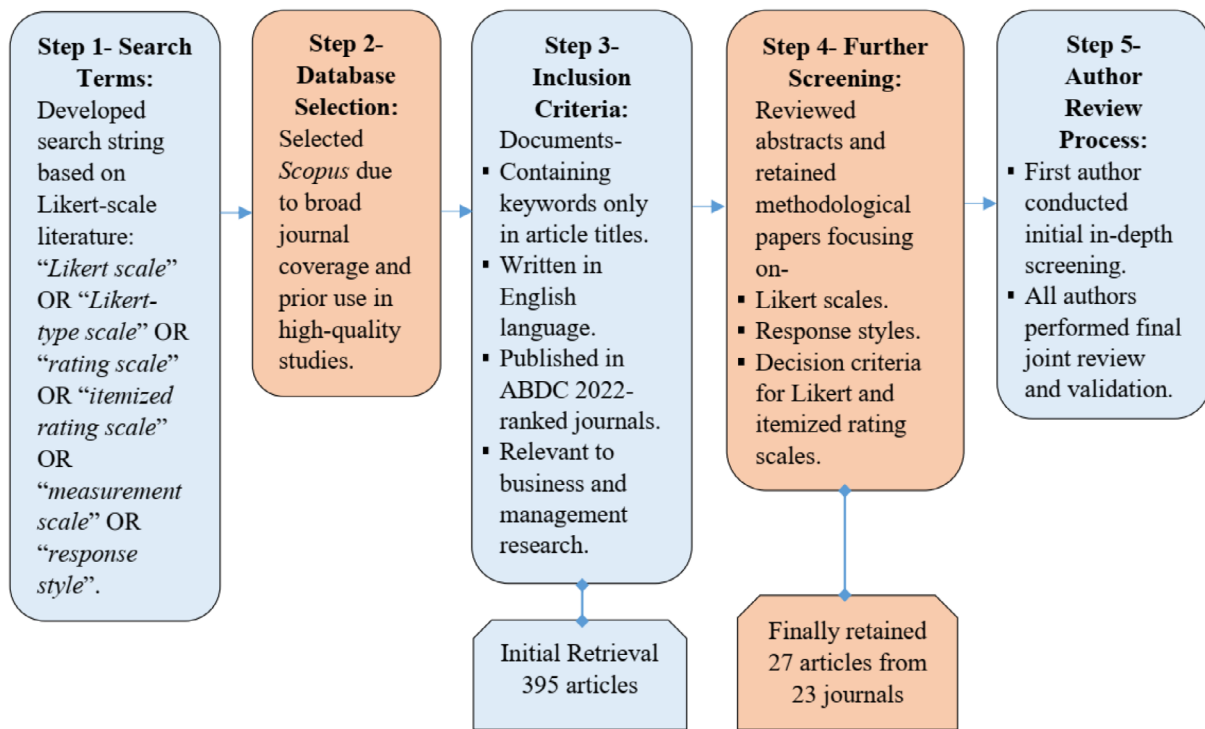


FIGURE 1 | The literature search, screening, and selection process.

3 | Major Decision Criteria

3.1 | Number of Scale Points

Since Likert's (1932) introduction of the Likert scale, researchers have extensively examined the optimal number of scale points to maximize reliability. However, no consensus has yet been reached on the ideal number of points to use. Some argued that reliability increases with the number of scale options (Finn et al. 2015; Hilbert et al. 2016), while others opined that reliability is not dependent on the number of scale points (Boote 1981; Brown et al. 1991; Jacoby and Matell 1971). Schutz and Rucker (1975) suggested that the number of categories on a scale does not essentially influence the cognitive structure derived from the results.

Former researchers suggested different numbers of scale points for maximum reliability, such as 7-point (Cicchetti et al. 1985; Eutsler and Lang 2015; Finn 1972; Matell and Jacoby 1972; Nunnally 1978) and 5-point (Chen et al. 2015; Jenkins and Taber 1976; Lissitz and Green 1975), while some studies even used a 4-point scale (e.g., Padgett and Morgan 2020). Employing a higher number of scale points may yield diminishing returns (Simms et al. 2019). For example, Bendig (1954) contended that reliability remains consistent with a 3-, 5-, 6-, or 9-point scale but diminishes with an 11-point scale. In line with this, Green and Rao (1970) observed that information retrieval is maximized by using 6- or 7-point response categories. They further noted that expanding the number of scale points beyond 7 might generate minimal additional information. An investigation of scale points (2–11) on a measure of service experience led by Preston and Colman (2000) confirmed earlier findings on diminishing returns. Specifically, the authors concluded that, when considering certain psychometrics (reliability, validity, and discriminating power), 2- to 4-point

scales perform poorly, while performance improves on scales of up to 7 points. Based on an examination with up to 6-point scales, Lee and Paek (2014) concluded that 4–6 is the perfect number of scale points. Indeed, 2- to 5-point scales were found to attenuate the psychometric advantage for response scales, with this ideal if extended up to 6 points, but not beyond (Simms et al. 2019). However, Harzing et al. (2009) illustrated that the 7-point rating scale yielded better results than the 5-point rating scale in cross-national studies.

The developers and/or the users of a scale should also take into consideration whether the scale points are specified as unipolar or bipolar (Jebb et al. 2021). For example, scale points ranging from “0” to “4” usually signify a unipolar scale, while “–3” to “+3” scale points imply a bipolar scale. The unipolarity or bipolarity of a scale is also implied in the verbal description of the scale point used. For example, labels, such as “not at all” to “extremely,” indicate unipolarity, whereas labeling scales ranging from “strongly disagree” to “strongly agree” imply bipolarity (Jebb et al. 2021).

The choice of the number of scale categories influences the response rate and midpoint response style (MRS) (as discussed in Section 3.2). To this end, Hartley and MacLean (2006) identified no significant differences in response rate with 3, 4, or 5 Likert-scale categories. However, a relatively large number of scale points (9 or 10) triggered a higher level of MRS, while, with a smaller number of scale categories, no MRS emerged (Kieruj and Moors 2010). In summary, no “one-size-fits-all” decision criterion exists for the number of scale points for an itemized rating scale. The decision should, therefore, be based on the study's context, in which cultural differences could be one of the major factors, as highlighted in Section 3.2.

3.2 | Use of Midpoints

Another important and often debated decision when using a Likert-type scale is whether to include a midpoint. A common concern with midpoints is the MRS, which reflects a tendency for respondents to disproportionately select the middle response category (Weijters 2006). The issue is even more critical when a midpoint could be offering an easy way out to many respondents (O’Muirheartaigh et al. 2000). From a rational standpoint, researchers typically favor respondents who make a definitive choice rather than opting for a neutral or intermediate position on a scale. Consequently, a scale without a midpoint would be preferable as long as it did not compromise the reliability or validity of the gathered information (Garland 1991). However, research findings on this issue have remained inconsistent.

A recent review has suggested that a rating scale of greater than 5 scale points (e.g., a 7-point scale) with an odd number of scale categories provides better reliability and validity coefficients (Kusmaryono et al. 2022). Worcester and Burns (1975) proposed that the absence of midpoints drives respondents more toward the positive end of the scale. Conversely, Garland (1991) opined that how people respond to a Likert-type scale without a midpoint is content-specific. He further suggested that eliminating the midpoint option in the scale can mitigate social desirability bias, wherein respondents may wish to please the inquirer, appear helpful, or avoid providing what is generally perceived as a socially undesirable answer. Besides, midpoints might cause ambiguity, proliferating measurement error if respondents perceive it differently, given that midpoints may not accurately represent their response (Simms et al. 2019). The reasons why respondents choose midpoints on an odd-numbered Likert-type scale include: (1) respondents actually have a modest response about the measured items, (2) respondents face complexity when choosing their opinion on the item, (3) respondents lack a proper understanding of the item, and (4) respondents perceive that their opinion is context-dependent (“it depends” thinking) (Kulas and Stachowski 2013). The latter three reasons are problematic as they fail to reflect respondents’ true perceptions or opinions, thereby increasing measurement error (Simms et al. 2019).

The inclusion of midpoints has also been examined in relation to scale reliability. Weems and Onwuegbuzie (2001) found a notable prevalence of midpoint selections in their samples, suggesting that this response pattern may undermine, rather than enhance, reliability. In line with this finding, Tsang (2012) indicated that midpoints are not necessary to benefit the internal consistency of measurements. Similarly, Chang (1994) disagreed with using reliability as a criterion to judge the merit of midpoints. This argument could be further strengthened by Cronbach’s note (Cronbach 1950, 22): “there is no merit in enhancing test reliability unless validity is enhanced at least proportionately.” In other words, validity should be a better criterion than reliability (Chang 1994).

The decision whether to include midpoints in Likert scales is also culture-specific (Harzing 2006). As Asian respondents, for example, have a tendency toward the middle response option (Beuthner et al. 2018; Harzing et al. 2012; Lee et al. 2002), many researchers in this region prefer to use an even number of response categories (see e.g., Ge et al. 2019; Li et al. 2019;

Taufique 2022). In one study on a group of Chinese, Japanese, and Americans, Lee et al. (2002) found that Chinese and Japanese respondents more frequently selected midpoints when presented with items that required admitting to a positive emotion. In another study, Chen et al. (1995) identified that respondents from two collectivist cultures (Chinese and Japanese) showed a greater preference for midpoints and a lesser inclination for extreme points than those from individualist cultures (North American). This finding was supported by another study (Grandy 1996), which revealed that Asian-American students are likelier to choose midpoints and avoid extreme points on a 5-point Likert scale in contrast to their white counterparts. Harzing et al. (2012) also concluded that Asian respondents have a higher tendency to choose middle response options. Likewise, Harumi (2011) revealed that respondents from Japan and Malaysia have higher average middle response tendencies than those of other cultures. These findings suggest that researchers should exercise caution in, or avoid, utilizing midpoints with specific justifications when dealing with respondents in an Asian context.

3.3 | Labeling of Response Categories

Labeling, also known as anchoring, refers to assigning a verbal description to scale points (e.g., 1 = “strongly disagree”). Endpoint labeling is common as it is easier to implement. For analyzing linear relationships such as correlations, regression, or structural equation modelling, 5- or 7-point scales with endpoint labeling are generally recommended (Weijters et al. 2010). However, reversed items (e.g., negatively framed items) may introduce bias in such formats (Weijters et al. 2010). To mitigate this, researchers should strategically disperse reversed items among buffer questions or alternate scales and include a method factor during data analysis (for details, see Weijters et al. 2009). On the other hand, labeling all response categories facilitates interpretation for researchers and respondents (Wildt and Mazis 1978). Fully labeled response categories are also linked to a higher level of reliability (Krosnick 1991; Alwin and Krosnick 1991). The intermediate options in fully labeled response categories tend to be more noticeable; thus, they are likely to attract an increased level of responses due to their high accessibility (Posavac et al. 1997, 2003). It is also argued that respondents consider the meaning of the labels provided with the response categories when choosing a specific response (Rohrmann 2003; Wildt and Mazis 1978). Hence, labeled response categories are believed to enhance the likelihood of attracting respondents (Garland 1990; Krosnick and Fabrigar 1997). Appendix A2 presents a list of different response anchors.

Researchers should also consider the effect of perceived intensity (e.g., the extremity of the measured items’ labeling) on response categories. Prior research, as detailed below, has found that label intensity was influential in response behavior. For instance, flexible absolute labels (e.g., “agree” and “disagree”) as anchors in the 5-point Likert scale resulted in a more even distribution of responses across all five scale categories, compared to strict absolute labels (e.g., “strongly agree” and “strongly disagree”), where responses were centered more on in-between scale categories (Wyatt and Meyers 1987). However, Lam and Stevens (1994) revealed that this intensity issue was linked to respondents’ understanding of the issue being measured and the structure

of item wording (e.g., “I do not buy it” vs. “I hate buying it”). Another critical issue pertains to whether respondents have different perceptions of the use of diverse amplifiers (e.g., “strongly” or “completely (dis)agree”) as endpoints in terms of intensity. In this regard, “extremely” was found to have a more intensifying impact than “very,” which again had more impact than “decidedly” (Cliff 1959). Similarly, Smith et al. (2009) reported differences in respondents’ perceived intensity across various amplifiers, such as “very much” and “definitely.”

Responses are influenced not only by perceived intensity but also by familiarity with scale points, particularly when surveys involve multiple languages requiring translation and back-translation. If an endpoint label in one language is more familiar than its counterpart in another, respondents are likely to select that endpoint more frequently in the former language (Weijters et al. 2013). This finding is particularly relevant for cross-national surveys where literal translations of the scale point labels may fall short, leading to idiomatic translations and label familiarity becoming critical concerns.

For example, the literal translation of “completely agree” in German is “*vollkommen einverstanden*,” and in Dutch, it is “*volledig eens*,” with the former scale point resulting in higher endpoint endorsements/responses due to its familiarity (in Germany and Switzerland) than the latter (in Belgium and the Netherlands) (Weijters et al. 2013). In this context, completely avoiding the amplifiers (e.g., using only “agree” and “disagree” as anchors) or labeling exclusively in an international language, such as English, are not considered viable solutions. Instead, careful attention should be given to labeling response categories in Likert-type scales to minimize cross-linguistic differences in familiarity across languages (Weijters et al. 2013). For example, extreme points such as “strongly” and “completely” (English version) can be literally translated to “*fortement*” and “*complètement*” in French. However, “very much” (English) and “*tout à fait*” (French) are somewhat language-specific expressions (Weijters et al. 2013) for which the exact meaning may not carry over across languages, and should therefore be avoided.

Labeling effectiveness can also be improved by incorporating relevant emojis (also known as ideograms or smileys) as response options, which may elicit stronger engagement in certain contexts, particularly among respondents with intellectual disabilities (Hartley and MacLean 2006). For example, a study involving individuals with mild to moderate intellectual disabilities found that response rates increased from 76% to 100% when scale options comprised smiling and frowning faces (Rojahn et al. 1994). Similarly, the use of emojis may be particularly relevant in high-context cultures, such as those in Asia, where non-verbal cues play an important role in communication (Cheng 2017).

3.4 | Presentation of Response Format (Horizontal vs. Vertical)

The use of online surveys has grown exponentially over the past decade (Ramsey et al. 2016; Zhang et al. 2017), particularly in social science research, enabling researchers to use different Likert-type scale formats (e.g., horizontal vs. vertical) to enhance respondents’ convenience and/or data quality (e.g., response rate)

(Liu 2017; Liu and Conrad 2016). However, responses generated from alternative scale formats have been found to vary, the result being poor internal consistency of measurement scales (Weijters et al. 2021). This is likely to be due to the vertically presented response categories being more prone to extreme response style (ERS) and the greater tendency to select endpoints, such as “strongly agree” or “strongly disagree,” compared to horizontally presented categories.

Weijters et al. (2021) also showed that visually compact response categories tend to produce higher ERS than visually spaced categories, even when both are labeled identically and presented horizontally. These authors named this phenomenon the “response category distance effect” (Weijters et al. 2021, 87). This effect is significant for online surveys where respondents respond using personal computers (PCs) or smartphones which may have different formats for response categories. Weijters et al. (2021) further recommended a few steps to mitigate the ERS effect: (i) a similar response device (a PC or a smartphone) should be ensured; (ii) the response format (either horizontal or vertical) should be identical for all measured items; and (iii) in the case of combining vertical and horizontal categories, the gap between (dropdown) categories (in the case of vertical) should be increased.

In summary, the critical decision criteria regarding the number of Likert scale points (e.g., 5-point vs. 7-point), use of midpoints, labeling of response categories, and presentation of response format (e.g., vertical vs. horizontal) are presented in Table 1, offering a quick overview for researchers employing the Likert scale in survey research.

4 | Implications for Researchers

The findings reported in this paper carry significant implications for researchers seeking to improve the quality and credibility of survey-based research using Likert-type scales. Many researchers have continued to adopt scale formats without fully considering their psychometric, cultural, and contextual implications. For instance, although 5-point and 7-point scales are dominant in the literature, their popularity often overshadows the nuanced trade-offs involved, such as the risk of reduced validity with too few options or cognitive overload with too many. The optimal number of points is not a universal rule, but rather a contextual decision that should reflect considerations of the scale’s reliability and validity, as well as the cognitive capacity of the respondent group. As discussed in Section 3.1, empirical evidence has suggested that 5-point and 7-point scales emerge as viable standards, offering a favorable trade-off between psychometric robustness and respondent comprehension (Clarke 2000; Weijters et al. 2010). Specifically, a 7-point scale may be more appropriate for samples with higher cognitive ability or familiarity with the subject matter, such as student populations, whereas a 5-point scale is often better suited for general population samples, where cognitive load and survey fatigue may pose greater concerns (Chen et al. 2015; Weijters et al. 2010). Moreover, the decision between unipolar (e.g., “0” to “4”) and bipolar (e.g., “+3” to “-3”) scale formats should be guided by the underlying nature of the construct being measured (Jebb et al. 2021). For instance, unipolar scales may be more suitable for measuring intensity-based constructs (e.g., satisfaction), while bipolar scales

TABLE 1 | Overview of major decision criteria.

Aspects	Findings
a) Number of scale points	<ul style="list-style-type: none"> ✓ No consensus on the optimal number of categories; 5-point and 7-point scales are commonly used (Weijters et al. 2010). ✓ Diminishing returns observed beyond 7 points (Simms et al. 2019). ✓ 6–7 points maximize reliability and information (Green and Rao 1970). ✓ Scales with fewer than 5 points may perform poorly in terms of reliability and validity (Preston and Colman 2000). ✓ Scale point variations, such as “0” to “4” (unipolar) and “+3” to “–3” (bipolar), and verbal descriptions, such as “not at all”—“extremely” (unipolar) and “strongly disagree”—“strongly agree” (bipolar), reflect distinct polarity continua in construct measurement (Jebb et al. 2021).
b) Use of midpoints	<ul style="list-style-type: none"> ✓ Midpoints may introduce ambiguity and measurement error when not accurately reflecting respondents’ opinions (Simms et al. 2019). ✓ Eliminating midpoints can reduce social desirability bias but may push responses toward the positive end of the scale (Garland 1991). ✓ Cultural tendencies, such as Asian respondents’ preference for midpoints (Harzing et al. 2012), suggest the avoidance or justification of midpoint usage in such contexts.
c) Labeling of response categories	<ul style="list-style-type: none"> ✓ Fully labeled categories improve interpretation and reliability (Alwin and Krosnick 1991). ✓ Labels should account for perceived intensity (e.g., “strongly” vs. “completely”) to ensure accurate measurement of responses (Cliff 1959; Smith et al. 2009). ✓ Cross-linguistic differences in label familiarity must be considered to avoid bias in international studies (Weijters et al. 2013). ✓ Emojis or ideograms can enhance engagement and comprehension, especially for respondents with limited literacy or intellectual disabilities (Hartley and MacLean 2006).
d) Presentation of response format	<ul style="list-style-type: none"> ✓ Horizontal formats generally reduce ERS compared to vertical formats (Weijters et al. 2021). ✓ Compact horizontal response categories may increase ERS; ensuring consistent formatting and spacing mitigates this effect (Weijters et al. 2021). ✓ Consistent formats across devices (e.g., PC or smartphone) enhance data quality in online surveys (Weijters et al. 2021).

may be better suited for assessing contrasting attitudes (e.g., preference).

Equally critical is the decision to include a midpoint. Research has suggested that midpoints can enhance data quality by reducing mis-responses to reverse-phrased items (Weijters et al. 2010), making them generally advisable unless specific contextual factors dictate otherwise. However, the assumption that midpoints function as neutral options is problematic, particularly in cross-cultural research contexts. For instance, in many Asian cultures, respondents demonstrate a well-documented tendency to select midpoints not as expressions of true neutrality, but rather as a culturally conditioned response style. This phenomenon has important methodological consequences: in these cultural contexts, omitting the midpoint may produce more discriminating and psychometrically sound data by preventing artificial central tendency bias. However, researchers should exercise caution when omitting midpoints, ensuring that the targeted respondents are sufficiently informed about the topic under investigation. One effective strategy is to use appropriate screening questions

to confirm their familiarity; alternatively, a “no response” option can be added to accommodate uncertainty.

Labeling and format choices are also highly significant. Endpoint labeling is easier to implement and is recommended for analyzing linear relationships, such as correlations, regression, and structural equation modeling (Weijters et al. 2010). Fully labeled scales may improve clarity and engagement, but they must be designed with sensitivity to linguistic nuances and cultural familiarity. This is especially relevant in multilingual or cross-national research, where subtle differences in phrasing can alter response patterns. Additionally, the physical layout of the scale (e.g., horizontal vs. vertical) can shape response styles in unexpected ways, particularly in digital surveys across varied devices. To mitigate these issues, researchers should pretest their surveys across multiple devices, browsers, and screen resolutions to identify and address potential formatting inconsistencies. Table 2 presents a summary of the key considerations and practical recommendations for researchers on the use of Likert scales in survey studies.

TABLE 2 | Summary of key considerations and practical recommendations for researchers.

Aspects	Implications/What to do
a) Number of scale points	<ul style="list-style-type: none"> ✓ For optimal results, use 7-point scales for student samples and 5-point scales for general population samples (Weijters et al. 2010). ✓ For intensity measures (e.g., satisfaction), use unipolar scales (“0” to “4”); for opposing attitudes (e.g., preference), bipolar scales (“+3” to “-3”) work better.
b) Use of midpoints	<ul style="list-style-type: none"> ✓ Use a midpoint to reduce mis-responses to reverse-worded questions. ✓ This decision could also vary across cultures. For example, for Asian respondents, exclude midpoints to prevent neutral bias and improve data quality.
c) Labeling of response categories	<ul style="list-style-type: none"> ✓ Use endpoint labeling for linear analyses (e.g., correlations, regression, structural equation modeling), but employ full labeling when developing scales (Weijters et al. 2010). However, when considering interpretation and reliability, full labeling is preferred. ✓ Select label intensity (e.g., “agree” vs. “strongly agree”), amplifiers (e.g., “very” vs. “extremely”), and phrasing based on the construct’s nature and respondents’ cultural context, guided by prior empirical evidence. ✓ Incorporate emojis when surveying respondents with intellectual disabilities or populations from high-context cultures (e.g., Asian contexts) to enhance comprehension and engagement.
d) Presentation of response format	<ul style="list-style-type: none"> ✓ Maintain consistent response formats (horizontal/vertical) and devices (PCs/smartphones) across all survey items to ensure data comparability (Weijters et al. 2021). ✓ Conduct pretests across multiple browsers and screen resolutions to identify and address potential formatting inconsistencies in online surveys (Weijters et al. 2021). ✓ Select items with means near the scale midpoint and use fully labeled response options to mitigate scale format effects.

5 | Concluding Remarks

These findings underscore the need to tailor the Likert-scale design to the research context, taking account of cultural, linguistic, and methodological factors. Clearly, no “one-size-fits-all” scale format is available for use in different research contexts, but informed decisions on scale points, midpoints, labeling of response categories, and appropriate presentation of response format can significantly enhance data quality and reliability. Researchers must justify their choices to ensure valid and meaningful results.

While this paper provides a concise overview of key considerations for Likert-scale design, it is not exhaustive. The findings are based on selected studies rather than a systematic literature review, which may limit the generalizability of the recommendations. Future research should conduct a comprehensive meta-analysis or systematic review to synthesize empirical evidence across disciplines, identifying domain-specific best practices.

Author Contributions

Khan Md Raziuddin Taufique: Conceptualization/idea, Literature search and data analysis, Writing – original draft preparation, Writing – review and editing, Supervision. **Md. Mahiuddin Sabbir:** Literature search and data analysis, Writing – original draft preparation, Writing – review and editing. **Fazlul K. Rabbane:** Literature search and data analysis, Writing – original draft preparation, Writing – review and editing, Supervision

Policy on Using ChatGPT and Similar AI Tools

The authors used ChatGPT and Copilot to check and improve the language (e.g., grammar, expression) of the paper. The authors reviewed the edits, made further edits, and take full responsibility for the content of the current paper.

Acknowledgments

Open access publishing facilitated by Curtin University, as part of the Wiley - Curtin University agreement via the Council of Australasian University Librarians

Ethics Statement

This study is based solely on previously published literature and does not involve human participants, animals, or primary data collection. Therefore, formal ethical approval was not required.

Consent

This submission does not contain any information that could identify any individual; therefore, ‘Consent to participate’ and/or ‘Consent to publish’ are not required.

Conflicts of Interest

None of the authors have a conflict of interest to disclose.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

Alabi, A. T., and M. O. Jelili. 2023. “Clarifying Likert Scale Misconceptions for Improved Application in Urban Studies.” *Quality and Quantity* 57, no. 2: 1337–1350. <https://doi.org/10.1007/s11135-022-01415-8>.

- AL-Abrow, H., M. Al-Maatoq, R. K. Alharbi, et al. 2021. "Understanding Employees' Responses to the COVID-19 Pandemic: The Attractiveness of Healthcare Jobs." *Global Business and Organizational Excellence* 40, no. 2: 19–33. <https://doi.org/10.1002/joe.22070>.
- Albaum, G. 1997. "The Likert Scale Revisited: An Alternate Version." *International Journal of Market Research* 39, no. 2: 1–21. <https://doi.org/10.1177/147078539703900202>.
- Alwin, D. F., and J. A. Krosnick. 1991. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes." *Sociological Methods & Research* 20, no. 1: 139–181. <https://doi.org/10.1177/0049124191020001005>.
- Baethge, C., S. Goldbeck-Wood, and S. Mertens. 2019. "SANRA—A Scale for the Quality Assessment of Narrative Review Articles." *Research Integrity and Peer Review* 4, no. 1: 5–11. <https://doi.org/10.1186/s41073-019-0064-8>.
- Bearden, W. O., and R. G. Netemeyer. 1999. *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*. Sage.
- Bendig, A. W. 1954. "Reliability and the Number of Rating-Scale Categories." *Journal of Applied Psychology* 38, no. 1: 38–40. <https://doi.org/10.1037/H0055647>.
- Benson, P. G., M. R. Buckley, and S. Hall. 1988. "The Impact of Rating Scale Format on Rater Accuracy: An Evaluation of the Mixed Standard Scale." *Journal of Management* 14, no. 3: 415–423. <https://doi.org/10.1177/014920638801400305>.
- Beuthner, C., M. Friedrich, C. Herbes, and I. Ramme. 2018. "Examining Survey Response Styles in Cross-Cultural Marketing Research: A Comparison Between Mexican and South Korean Respondents." *International Journal of Market Research* 60, no. 3: 257–267. <https://doi.org/10.1177/1470785318762015>.
- Boote, A. S. 1981. "Reliability Testing of Psychographic Scales: Five-Point or Seven-Point? Anchored or Labeled?" *Journal of Advertising Research* 2: 53–60.
- Brown, G., R. E. Widing, and R. L. Coulter. 1991. "Customer Evaluation of Retail Salespeople Utilizing the SOCO Scale: A Replication, Extension, and Application." *Journal of the Academy of Marketing Science* 19, no. 4: 347–351. <https://doi.org/10.1007/BF02726510>.
- Bruner, G. C., P. J. Hensel, and K. E. James. 2001. *Marketing Scales Handbook: A Compilation of Multi-Item Measures*. 3rd ed. American Marketing Association.
- 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, no. 3: 205–215. <https://doi.org/10.1177/014662169401800302>.
- Chen, C., S. Y. Lee, and H. W. Stevenson. 1995. "Response Style and Cross-Cultural Comparisons of Rating Scales Among East Asian and North American Students." *Psychological Science* 6, no. 3: 170–175. <https://doi.org/10.1111/j.1467-9280.1995.tb00327.x>.
- Chen, X., H. Yu, and F. Yu. 2015. "What Is the Optimal Number of Response Alternatives for Rating Scales? From an Information Processing Perspective." *Journal of Marketing Analytics* 3, no. 2: 69–78. <https://doi.org/10.1057/jma.2015.4>.
- Cheng, L. 2017. "Do I Mean What I Say and Say What I Mean? A Cross Cultural Approach to the Use of Emoticons & Emojis in CMC Messages." *Fonseca, Journal of Communication* 15: 207–226. <https://doi.org/10.14201/fjc201715199217>.
- Chyung, S. Y. Y., K. Roberts, I. Swanson, and A. Hankinson. 2017. "Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale." *Performance Improvement* 56, no. 10: 15–23. <https://doi.org/10.1002/pfi.21727>.
- Cicchetti, D. V., D. Shoinralter, and P. J. Tyrer. 1985. "The Effect of Number of Rating Scale Categories on Levels of Interrater Reliability: A Monte Carlo Investigation." *Applied Psychological Measurement* 9, no. 1: 31–36. <https://doi.org/10.1177/014662168500900103>.
- Clarke, I. 2000. "Global Marketing Research: Is Extreme Response Style Influencing Your Results?" *Journal of International Consumer Marketing* 12, no. 4: 91–111. https://doi.org/10.1300/J046v12n04_06.
- Cliff, N. 1959. "Adverbs as Multipliers." *Psychological Review* 66, no. 1: 27–44. <https://doi.org/10.1037/h0045660>.
- Cronbach, L. J. 1950. "Further Evidence on Response Sets and Test Design." *Educational and Psychological Measurement* 10, no. 1: 3–31. <https://doi.org/10.1177/001316445001000101>.
- Demiris, G., D. P. Oliver, and K. T. Washington. 2019. "The Foundations of Behavioral Intervention Research in Hospice and Palliative care." In *Behavioral Intervention Research in Hospice and Palliative Care: Building an Evidence Base*, 17–25. Academic Press. <https://doi.org/10.1016/B978-0-12-814449-7.00002-8>.
- Dhaliwal, A., D. P. Singh, and J. Paul. 2025. "The Consumer Behavior of Luxury Goods: A Review and Research Agenda." *Journal of Strategic Marketing* 33, no. 1: 66–92. <https://doi.org/10.1080/0965254X.2020.1758198>.
- Dolnicar, S. 2021. "5/7-Point 'Likert Scales' Aren't Always the Best Option." *Annals of Tourism Research* 91: 103297. <https://doi.org/10.1016/j.annals.2021.103297>.
- Eutsler, J., and B. Lang. 2015. "Rating Scales in Accounting Research: The Impact of Scale Points and Labels." *Behavioral Research in Accounting* 27, no. 2: 35–51. <https://doi.org/10.2308/bria-51219>.
- Finn, J. A., Y. S. Ben-Porath, and A. Tellegen. 2015. "Dichotomous Versus Polytomous Response Options in Psychopathology Assessment: Method or Meaningful Variance?" *Psychological Assessment* 27, no. 1: 184–193. <https://doi.org/10.1037/PAS0000044>.
- Finn, R. H. 1972. "Effects of Some Variations in Rating Scale Characteristics on the Means and Reliabilities of Ratings." *Educational and Psychological Measurement* 32, no. 2: 255–265. <https://doi.org/10.1177/001316447203200203>.
- Garland, R. 1990. "A Comparison of Three Forms of the Semantic Differential." *Marketing Bulletin* 1, no. 1: 19–24.
- Garland, R. 1991. "The Mid-Point on a Rating Scale: Is It Desirable?" *Marketing Bulletin* 2, no. 1: 66–70.
- Ge, J., M. Carney, and F. Kellermanns. 2019. "Who Fills Institutional Voids? Entrepreneurs' Utilization of Political and Family Ties in Emerging Markets." *Entrepreneurship: Theory and Practice* 43, no. 6: 1124–1147. <https://doi.org/10.1177/1042258718773175>.
- Gomes, P., E. Santos, and E. Martins. 2023. "An Exploratory Analysis of Internal Communication in Times of the COVID-19 Pandemic." *Global Business and Organizational Excellence* 42, no. 5: 37–49. <https://doi.org/10.1002/joe.22204>.
- Grandy, J. 1996. "Differences in the Survey Responses of Asian American and White Science and Engineering Students." In *Educational Testing Service (ETS) Research Report Series* 1996, no. 2: 1–23. <https://doi.org/10.1002/j.2333-8504.1996.tb01703.x>.
- Green, P. E., and V. R. Rao. 1970. "Rating Scales and Information Recovery—How Many Scales and Response Categories to Use?" *Journal of Marketing* 34, no. 3: 33–39. <https://doi.org/10.1177/002224297003400307>.
- Gupta, D. G., H. Shin, and V. Jain. 2023. "Luxury Experience and Consumer Behavior: A Literature Review." *Marketing Intelligence and Planning* 41, no. 2: 199–213. <https://doi.org/10.1108/MIP-12-2021-0438>.
- Hamamura, T., S. J. Heine, and D. L. Paulhus. 2008. "Cultural Differences in Response Styles: The Role of Dialectical Thinking." *Personality and Individual Differences* 44, no. 4: 932–942. <https://doi.org/10.1016/j.paid.2007.10.034>.
- Haque, M. A., M. A. Islam, and S. H. Soh. 2024. "Green HRM Practices and Green Work Engagement: The Roles of Green Knowledge Sharing and Green Group Cohesion." *Global Business and Organizational Excellence* 44, no. 1: 16–31. <https://doi.org/10.1002/joe.22266>.

- Hartley, S. L., and W. E. MacLean Jr. 2006. "A Review of the Reliability and Validity of Likert-Type Scales for People With Intellectual Disability." *Journal of Intellectual Disability Research* 50, no. 11: 813–827. <https://doi.org/10.1111/j.1365-2788.2006.00844.x>.
- Harumi, C. A. 2011. "Cross-Cultural Differences in Response Styles." Doctoral diss., Washington State University.
- Harzing, A. W. 2006. "Response Styles in Cross-National Survey Research: A 26-Country Study." *International Journal of Cross Cultural Management* 6, no. 2: 243–266. <https://doi.org/10.1177/1470595806066332>.
- Harzing, A. W., J. Balduenza, W. Barner-Rasmussen, et al. 2009. "Rating Versus Ranking: What Is the Best Way to Reduce Response and Language Bias in Cross-National Research?" *International Business Review* 18, no. 4: 417–432. <https://doi.org/10.1016/j.ibusrev.2009.03.001>.
- Harzing, A. W., M. Brown, K. Köster, and S. Zhao. 2012. "Response Style Differences in Cross-National Research: Dispositional and Situational Determinants." *Management International Review* 52, no. 3: 341–363. <https://doi.org/10.1007/s11575-011-0111-2>.
- Heo, C. Y., B. Kim, K. Park, and R. M. Back. 2022. "A Comparison of Best-Worst Scaling and Likert Scale Methods on Peer-to-Peer Accommodation Attributes." *Journal of Business Research* 148: 368–377. <https://doi.org/10.1016/j.jbusres.2022.04.064>.
- Hilbert, S., H. Küchenhoff, N. Sarubin, T. T. Nakagawa, and M. Bühner. 2016. "The Influence of the Response Format in a Personality Questionnaire: An Analysis of a Dichotomous, a Likert-Type, and a Visual Analogue Scale." *TPM—Testing, Psychometrics, Methodology in Applied Psychology* 23, no. 1: 3–24. <https://doi.org/10.4473/tpm23.1.1>.
- Hossain, M. A. 2019. "Effects of Uses and Gratifications on Social Media Use: The Facebook Case With Multiple Mediator Analysis." *PSU Research Review* 3, no. 1: 16–28. <https://doi.org/10.1108/prr-07-2018-0023>.
- Jacoby, J., and M. S. Matell. 1971. "Three-Point Likert Scales Are Good Enough." *Journal of Marketing Research* 8, no. 4: 495–500. <https://doi.org/10.1177/002224377100800414>.
- Jebb, A. T., V. Ng, and L. T. Tay. 2021. "A Review of Key Likert Scale Development Advances: 1995–2019." *Frontiers in Psychology* 12: 637547. <https://doi.org/10.3389/fpsyg.2021.637547>.
- Jenkins, G. D., and T. D. Taber. 1976. "A Monte Carlo Study of Factors Affecting Three Indices of Composite Scale Reliability." *Journal of Applied Psychology* 62, no. 4: 392–398. <https://doi.org/10.1037/0021-9010.62.4.392>.
- Kieruj, N. D., and G. Moors. 2010. "Variations in Response Style Behavior by Response Scale Format in Attitude Research." *International Journal of Public Opinion Research* 22, no. 3: 320–342. <https://doi.org/10.1093/ijpor/edq001>.
- Krosnick, J. A. 1991. "Response Strategies for Coping With the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5, no. 3: 213–236. <https://doi.org/10.1002/ACP.2350050305>.
- Krosnick, J. A., and L. R. Fabrigar. 1997. "Designing Rating Scales for Effective Measurement in Surveys." In *Survey Measurement and Process*, edited by L. E. Lyberg, P. Biemer, M. Collins, E. D. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin, 141–164. John Wiley & Sons.
- Kulas, J. T., and A. A. Stachowski. 2013. "Respondent Rationale for Neither Agreeing nor Disagreeing: Person and Item Contributors to Middle Category Endorsement Intent on Likert Personality Indicators." *Journal of Research in Personality* 47, no. 4: 254–262. <https://doi.org/10.1016/j.jrmp.2013.01.014>.
- Kumar, M., K. A. Rezo, and J. A. Karim. 2008. "Reexamining the Consistency of Likert Type Scale in Primary Research." *International Journal of Operations and Quantitative Management* 14, no. 1: 61–70.
- Kusmaryono, I., D. Wijayanti, and H. R. Maharani. 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, no. 4: 625–637. <https://doi.org/10.12973/ijem.8.4.625>.
- Lam, T. C. M., and J. J. Stevens. 1994. "Effects of Content Polarization, Item Wording, and Rating Scale Width on Rating Response." *Applied Measurement in Education* 7, no. 2: 141–158. https://doi.org/10.1207/s15324818ame0702_3.
- Lee, J., and I. Paek. 2014. "In Search of the Optimal Number of Response Categories in a Rating Scale." *Journal of Psychoeducational Assessment* 32, no. 7: 663–673. <https://doi.org/10.1177/0734282914522200>.
- Lee, J. W., P. S. Jones, Y. Mineyama, and X. E. Zhang. 2002. "Cultural Differences in Responses to a Likert Scale." *Research in Nursing & Health* 25, no. 4: 295–306. <https://doi.org/10.1002/nur.10041>.
- Li, Q. 2013. "A Novel Likert Scale Based on Fuzzy Sets Theory." *Expert Systems with Applications* 40, no. 5: 1609–1618. <https://doi.org/10.1016/j.eswa.2012.09.015>.
- Li, Q., Q. Gu, and W. He. 2019. "Resilience of Chinese Teachers: Why Perceived Work Conditions and Relational Trust Matter." *Measurement* 17, no. 3: 143–159. <https://doi.org/10.1080/15366367.2019.1588593>.
- Likert, R. 1932. "A Technique for the Measurement of Attitudes." *Archives of Psychology* 22, no. 140: 1–55.
- Lim, T. Y., B. C. Y. Lim, C. M. Leong, I. G. Phang, and W. H. Foong. 2023. "Consumer Adoption of On-Demand Digital Platforms: An Integrated Model." *Global Business and Organizational Excellence* 42, no. 6: 75–88. <https://doi.org/10.1002/joc.22210>.
- Lim, W. M., K. M. Jasim, and A. Malathi. 2025. "Service Robots in Healthcare: Toward a Healthcare Service Robot Acceptance Model (sRAM)." *Technology in Society* 82, no. April: 102932. <https://doi.org/10.1016/j.techsoc.2025.102932>.
- Lissitz, R. W., and S. B. Green. 1975. "Effect of the Number of Scale Points on Reliability: A Monte Carlo Approach." *Journal of Applied Psychology* 60, no. 1: 10–13. <https://doi.org/10.1037/H0076268>.
- Liu, M. 2017. "Web Survey Experiments on Matrix Questions." *Computers in Human Behavior* 67: 61–72. <https://doi.org/10.1016/j.chb.2016.10.017>.
- Liu, M., and F. G. Conrad. 2016. "An Experiment Testing Six Formats of 101-Point Rating Scales." *Computers in Human Behavior* 55: 364–371. <https://doi.org/10.1016/j.chb.2015.09.036>.
- Makanyeza, C., B. K. Mabenge, and G. P. K. Ngorora-Madzimure. 2023. "Factors Influencing Small and Medium Enterprises' Innovativeness: Evidence From Manufacturing Companies in Harare, Zimbabwe." *Global Business and Organizational Excellence* 42, no. 3: 10–23. <https://doi.org/10.1002/joc.22180>.
- Matell, M. S., and J. Jacoby. 1971. "Is There an Optimal Number of Alternatives for Likert Scale Items? Study 1: Reliability and Validity." *Educational and Psychological Measurement* 31, no. 3: 657–674. <https://doi.org/10.1177/001316447103100307>.
- Matell, M. S., and J. Jacoby. 1972. "Is There an Optimal Number of Alternatives for Likert-Scale Items? Effects of Testing Time and Scale Properties." *Journal of Applied Psychology* 56, no. 6: 506–509. <https://doi.org/10.1037/h0033601>.
- Maulana, H., N. Khawaja, and P. Obst. 2019. "Development and Validation of the Indonesian Well-Being Scale." *Asian Journal of Social Psychology* 22, no. 3: 268–280.
- Miyatake, S., and M. Higuchi. 2017. "Does Religious Priming Increase the Prosocial Behaviour of a Japanese Sample in an Anonymous Economic Game?" *Asian Journal of Social Psychology* 20, no. 1: 54–59. <https://doi.org/10.1111/ajsp.12164>.
- Nørfelt, A., F. Kock, and S. Filep. 2025. "Tourism Psychology: Past, Present, and Future." *Annals of Tourism Research* 115: 104040. <https://doi.org/10.1016/j.annals.2025.104040>.
- Nunnally, J. C. 1978. *Psychometric Theory*, 2nd ed. McGraw-Hill.
- Oliveira, M., J. R. Andrade, V. Ratten, and E. Santos. 2023. "Psychological Empowerment for the Future of Work: Evidence From Portugal." *Global Business and Organizational Excellence* 42, no. 5: 65–78. <https://doi.org/10.1002/joc.22194>.

- O'Muircheartaigh, C., J. Krosnick, and A. Helic. 2000. "Middle Alternatives, Acquiescence, and the Quality of Questionnaire Data." Working paper, Harris School of Public Policy Studies, University of Chicago. https://www.researchgate.net/publication/5091207_Middle_Alternatives_Acquiescence_and_the_Quality_of_Questionnaire_Data/citations.
- Padgett, R. N., and G. B. Morgan. 2020. "Using the eRm Package for Rasch Modeling." *Measurement* 18, no. 3: 163–176. <https://doi.org/10.1080/15366367.2020.1732155>.
- Paramitha, V., S. Z. Tan, and W. M. Lim. 2025. "Undoing Greenwashing: The Roles of Greenwashing Severity, Consumer Forgiveness, Growth Beliefs and Apology Sincerity." *Journal of Product and Brand Management* 34, no. 4: 433–449. <https://doi.org/10.1108/jpbm-11-2023-4846>.
- Podsakoff, P. M., N. P. Podsakoff, L. J. Williams, C. Huang, and J. Yang. 2024. "Common Method Bias: It's Bad, It's Complex, It's Widespread, and It's Not Easy to Fix." *Annual Review of Organizational Psychology and Organizational Behavior* 11: 17–61. <https://doi.org/10.1146/annurev-orgpsych-110721-040030>.
- Posavac, S. S., M. Herzenstein, and D. M. Sanbonmatsu. 2003. "The Role of Decision Importance and the Salience of Alternatives in Determining the Consistency Between Consumers' Attitudes and Decisions." *Marketing Letters* 14, no. 1: 47–57. <https://doi.org/10.1023/A:1022858118201>.
- Posavac, S. S., D. M. Sanbonmatsu, and R. H. Fazio. 1997. "Considering the Best Choice: Effects of the Salience and Accessibility of Alternatives on Attitude–Decision Consistency." *Journal of Personality and Social Psychology* 72, no. 2: 253–261. <https://doi.org/10.1037/0022-3514.72.2.253>.
- Presbitero, A., Y. Fujimoto, and W. M. Lim. 2025. "Employee Engagement and Retention in Multicultural Work Groups: The Interplay of Employee and Supervisory Cultural Intelligence." *Journal of Business Research* 186. <https://doi.org/10.1016/j.jbusres.2024.115012>.
- Preston, C. C., and A. M. Colman. 2000. "Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences." *Acta Psychologica* 104, no. 1: 1–15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5).
- Ramsey, S. R., K. L. Thompson, M. McKenzie, and A. Rosenbaum. 2016. "Psychological Research in the Internet Age: The Quality of Web-Based Data." *Computers in Human Behavior* 58: 354–360. <https://doi.org/10.1016/j.chb.2015.12.049>.
- Rocereto, J. F., M. Puzakova, R. E. Anderson, and H. Kwak. 2011. "The Role of Response Formats on Extreme Response Style: A Case of Likert-Type vs. Semantic Differential Scales." *Advances in International Marketing* 22: 53–71. [https://doi.org/10.1108/S1474-7979\(2011\)0000022006](https://doi.org/10.1108/S1474-7979(2011)0000022006).
- Rohrmann, B. 2003. *Verbal Qualifiers for Rating Scales: Sociolinguistic Considerations and Psychometric Data*. <http://www.rohrmannresearch.net/pdfs/rohrmann-vqs-report.pdf>.
- Rojahn, J., V. J. Warren, and S. Ohringer. 1994. "A Comparison of Assessment Methods for Depression in Mental Retardation." *Journal of Autism and Developmental Disorders* 24, no. 3: 305–313. <https://doi.org/10.1007/BF02172229>.
- Russell, C. J., and P. Bobko. 1992. "Moderated Regression Analysis and Likert Scales: Too Coarse for Comfort." *Journal of Applied Psychology* 77, no. 3: 336–342. <https://doi.org/10.1037/0021-9010.77.3.336>.
- Sabbir, M. M., T. T. Khan, A. Das, S. Akter, and M. A. Hossain. 2023. "Understanding the Determinants of Consumers' Reverse Exchange Intention as an Approach to E-Waste Recycling: A Developing Country Perspective." *Asia-Pacific Journal of Business Administration* 15, no. 3: 411–439. <https://doi.org/10.1108/APJBA-11-2021-0565>.
- Sabbir, M. M., K. M. R. Taufique, and M. Nomi. 2021. "Telemedicine Acceptance During the COVID-19 Pandemic: User Satisfaction and Strategic Healthcare Marketing Considerations." *Health Marketing Quarterly* 38, no. 2–3: 168–187. <https://doi.org/10.1080/07359683.2021.1986988>.
- Sarah, F. H., C. L. Goi, F. Chieng, and K. M. R. Taufique. 2021. "Examining the Influence of Atmospheric Cues on Online Impulse Buying Behavior Across Product Categories: Insights From an Emerging E-market." *Journal of Internet Commerce* 20, no. 1: 25–45. <https://doi.org/10.1080/15332861.2020.1836593>.
- Schutz, H. G., and M. H. Rucker. 1975. "A Comparison of Variable Configurations Across Scale Lengths: An Empirical Study." *Educational and Psychological Measurement* 35, no. 2: 319–324. <https://doi.org/10.1177/001316447503500210>.
- ShabbirHusain, R. V., R. Kamath, and J. Moorthy. 2025. "Navigating the Sustainability–Luxury Paradox: Bibliometric Insights and Research Directions." *Journal of Consumer Behaviour* 24, no. 3: 1222–1248. <https://doi.org/10.1002/cb.2469>.
- Simms, L. J., K. Zelazny, T. F. Williams, and L. Bernstein. 2019. "Does the Number of Response Options Matter? Psychometric Perspectives Using Personality Questionnaire Data." *Psychological Assessment* 31, no. 4: 557–566. <https://doi.org/10.1037/pas0000648>.
- Smith, T. W., P. P. Mohler, J. Harkness, and N. Onodera. 2009. "Methods for Assessing and Calibrating Response Scales Across Countries and Languages." In *New Frontiers in Comparative Sociology*, edited by M. Sasaki, 45–79. Brill. <https://doi.org/10.1163/156913305775010106>.
- Soon, P. S., and W. M. Lim. 2025. "Consumer Emotions and Personality Traits in Augmented Reality." *Journal of Consumer Behaviour* 24, no. 5: 2214–2250. <https://doi.org/10.1002/cb.70000>.
- Sukhera, J. 2022. "Narrative Reviews: Flexible, Rigorous, and Practical." *Journal of Graduate Medical Education* 14, no. 4: 414–417. <https://doi.org/10.4300/JGME-D-22-00480.1>.
- Taufique, K. M. R. 2022. "Integrating Environmental Values and Emotion in Green Marketing Communications Inducing Sustainable Consumer Behaviour." *Journal of Marketing Communications* 28, no. 3: 272–290. <https://doi.org/10.1080/13527266.2020.1866645>.
- Taufique, K. M. R., M. M. Sabbir, S. Quinton, and S. S. Andaleeb. 2024. "The Different Impact of Utilitarian and Hedonic Attributes on Web-Based Retail Shopping Behaviour Through the Lens of Extended Technology Acceptance Model." *International Journal of Retail and Distribution Management* 52, no. 4: 443–460. <https://doi.org/10.1108/IJRDM-08-2023-0505>.
- Taufique, K. M. R., A. Vocino, and M. J. Polonsky. 2017. "The Influence of Eco-Label Knowledge and Trust on Pro-Environmental Consumer Behaviour in an Emerging Market." *Journal of Strategic Marketing* 25, no. 7: 511–529. <https://doi.org/10.1080/0965254X.2016.1240219>.
- Tortajada, E. G., A. C. Moreira, P. Duarte, and S. C. Silva. 2024. "Circular Economy and Sustainability in Luxury Fashion Consumer Behavior: A Review and Research Agenda." *International Journal of Consumer Studies* 48, no. 5: e13089. <https://doi.org/10.1111/ijcs.13089>.
- Tsang, K. K. 2012. "The Use of Midpoint on Likert Scale: The Implications for Educational Research." *Hong Kong Teachers' Centre Journal* 11, no. 1: 121–130.
- Tuan, T., T. Harpham, N. T. Huong, et al. 2005. "Validity of a Social Capital Measurement Tool in Vietnam." *Asian Journal of Social Science* 33, no. 2: 208–222. <https://www.jstor.org/stable/23654359>.
- van Dinther, M., F. Dochy, and M. Segers. 2011. "Factors Affecting Students' Self-Efficacy in Higher Education." *Educational Research Review* 6, no. 2: 95–108. <https://doi.org/10.1016/j.edurev.2010.10.003>.
- Wakita, T., N. Ueshima, and H. Noguchi. 2012. "Psychological Distance Between Categories in the Likert Scale: Comparing Different Numbers of Options." *Educational and Psychological Measurement* 72, no. 4: 533–546. <https://doi.org/10.1177/0013164411431162>.
- Weems, G. H., and A. J. Onwuegbuzie. 2001. "The Impact of Midpoint Responses and Reverse Coding on Survey Data." *Measurement and Evaluation in Counseling and Development* 34, no. 3: 166–176. <https://doi.org/10.1080/07481756.2002.12069033>.
- Weijters, B. 2006. Response Styles in Consumer Research. Ghent University. Faculty of Economics and Business Administration, Ghent, Belgium. <https://biblio.ugent.be/publication/4100284/file/4100290.pdf>.

- Weijters, B., E. Cabooter, and N. Schillewaert. 2010. "The Effect of Rating Scale Format on Response Styles: The Number of Response Categories and Response Category Labels." *International Journal of Research in Marketing* 27, no. 3: 236–247. <https://doi.org/10.1016/j.ijresmar.2010.02.004>.
- Weijters, B., M. Geuens, and H. Baumgartner. 2013. "The Effect of Familiarity With the Response Category Labels on Item Response to Likert Scales." *Journal of Consumer Research* 40, no. 2: 368–381. <https://doi.org/10.1086/670394>.
- Weijters, B., M. Geuens, and N. Schillewaert. 2009. "The Proximity Effect: The Role of Inter-Item Distance on Reverse-Item Bias." *International Journal of Research in Marketing* 26, no. 1: 2–12. <https://doi.org/10.1016/j.ijresmar.2008.09.003>.
- Weijters, B., K. Millet, and E. Cabooter. 2021. "Extremity in Horizontal and Vertical Likert Scale Format Responses. Some Evidence on How Visual Distance Between Response Categories Influences Extreme Responding." *International Journal of Research in Marketing* 38, no. 1: 85–103. <https://doi.org/10.1016/j.ijresmar.2020.04.002>.
- Weijters, B., N. Schillewaert, and M. Geuens. 2008. "Assessing Response Styles Across Modes of Data Collection." *Journal of the Academy of Marketing Science* 36, no. 3: 409–422. <https://doi.org/10.1007/s11747-007-0077-6>.
- Wildt, A. R., and M. B. Mazis. 1978. "Determinants of Scale Response: Label Versus Position." *Journal of Marketing Research* 15, no. 2: 261–267. <https://doi.org/10.1177/002224377801500209>.
- Worcester, R. M., and T. R. Burns. 1975. "A Statistical Examination of the Relative Precision of Verbal Scales." *Journal of the Market Research Society* 17, no. 3: 181–197.
- Wyatt, R. C., and L. S. Meyers. 1987. "Psychometric Properties of Four 5-Point Likert Type Response Scales." *Educational and Psychological Measurement* 47, no. 1: 27–35. <https://doi.org/10.1177/0013164487471003>.
- Yadav, S. S., and S. K. Kar. 2024. "Consumers' Purchase Intention of Private Labels: The Case of E-Retailers." *Global Business and Organizational Excellence* 44, no. 1: 47–64. <https://doi.org/10.1002/joe.22268>.
- Zhang, X. C., L. Kuchinke, M. L. Woud, J. Velten, and J. Margraf. 2017. "Survey Method Matters: Online/Offline Questionnaires and Face-to-Face or Telephone Interviews Differ." *Computers in Human Behavior* 71: 172–180. <https://doi.org/10.1016/J.CHB.2017.02.006>.

APPENDIX

A1. Summary of the reviewed papers

Authors	Journal name	ABDC category	Title	Key findings
Green and Rao (1970)	<i>Journal of Marketing</i>	A*	Rating scales and information recovery—How many scales and response categories to use?	Rating scales with about 6–7 response categories and roughly 8 well-balanced scales provide optimal information recovery, whereas coarse 2–3-point scales significantly distort results and cannot be compensated for by adding more scales.
Jacoby and Matell (1971)	<i>Journal of Marketing Research</i>	A*	Three-point Likert scales are good enough	Reliability and validity of Likert-type scales do not depend on the number of response points; 3-point scales are sufficient, as most variance comes from response direction rather than intensity.
Matell and Jacoby (1971)	<i>Educational and Psychological Measurement</i>	A	Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and Validity	Likert scale reliability and validity are largely independent of the number of response points, meaning 2- or 3-point scales can perform as well as multi-step scales. While early studies suggested 7–9 points for optimal measurement, too few points reduce discrimination, and too many exceed rater ability, so no single “best” number exists.
Matell and Jacoby (1972)	<i>Journal of Applied Psychology</i>	A*	Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties	Increasing the number of Likert-scale response options does not improve reliability, validity, or the proportion of scale used, while testing time increases only slightly. However, at larger scales, neutral or “uncertain” responses are significantly reduced, allowing more precise expression of attitudes.
Benson et al. (1988)	<i>Journal of Management</i>	A*	The impact of rating scale format on rater accuracy: An evaluation of the mixed standard scale	Behaviorally anchored rating scales [i.e., qualitative narratives with numerical ratings (e.g., 1–5)] produce more accurate performance ratings than mixed standard scale [e.g., better than (+), equal to (0)] formats.
Garland (1991)	<i>Marketing Bulletin</i>	C	The mid-point on a rating scale: Is it desirable?	Social desirability bias—stemming from respondents’ desire to please the interviewer or appear helpful—can be reduced by removing the midpoint category (e.g., “neither . . . nor” or “uncertain”) from Likert scales.
Russell and Bobko (1992)	<i>Journal of Applied Psychology</i>	A*	Moderated regression analysis and Likert scales: Too coarse for comfort	Using coarse Likert scales reduces the detection of true moderator effects, leading to information loss, whereas continuous scales increase effect sizes by ~93%, highlighting the need for more precise measurement.

(Continues)

Authors	Journal name	ABDC category	Title	Key findings
Chang (1994)	<i>Applied Psychological Measurement</i>	B	A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity	The choice between 4- and 6-point scales may not have a universally optimal answer and is likely to depend on the specific empirical context.
Albaum (1997)	<i>International Journal of Market Research</i>	A	The Likert scale revisited: An alternate version	Test of predictive ability showed that the two-stage format was a better predictor of product preferences than the one-stage format.
Clarke (2000)	<i>Journal of International Consumer Marketing</i>	B	Global marketing research: Is extreme response style influencing your results?	Expanding Likert scale response options from 3–4 points to 5–7 points notably reduces ERS, with a 7-point scale offering the most effective balance.
Preston and Colman (2000)	<i>Acta Psychologica</i>	A	Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences	While the internal consistency of the scales was similar, scales with over 10 response options showed lower reliability over time. Respondents also clearly preferred the 10-point scale, with the 7- and 9-point scales being a close second.
Harzing (2006)	<i>International Journal of Cross Cultural Management</i>	B	Response styles in cross-national survey research: A 26-country study	Country-level traits—such as power distance, collectivism, uncertainty avoidance, and extraversion—significantly shape response styles, including acquiescence and extreme responding.
Hamamura et al. (2008)	<i>Personality and Individual Differences</i>	A	Cultural differences in response styles: The role of dialectical thinking	East Asian respondents showed more moderate and ambivalent survey responses than European-heritage respondents, and these cultural differences were largely explained by higher levels of dialectical thinking rather than actual differences in trait levels.
Kumar et al. (2008)	<i>International Journal of Operations and Quantitative Management</i>	C	Reexamining the consistency of Likert type scale in primary research	In the Malaysian context, respondents' answers changed significantly when the scale shifted from fewer to more points or from balanced to unbalanced formats. However, changing the scale from neutral to non-neutral did not produce significant differences in responses.

(Continues)

Authors	Journal name	ABDC category	Title	Key findings
Weijters et al. (2008)	<i>Journal of the Academy of Marketing Science</i>	A*	Assessing response styles across modes of data collection	Response style biases, like acquiescence and midpoint responding, can distort survey results across modes. Measurement invariance alone could not detect this, but the representative indicators' response style means, and covariance structure (RIRSMACS) effectively corrected it, showing that apparent differences were due to response styles, not real content.
Kieruj and Moors (2010)	<i>International Journal of Public Opinion Research</i>	A	Variations in response style behavior by response scale format in attitude research	Extreme response style (ERS) is unaffected by the number of scale points, whereas the tendency to use the midpoint response emerges only in longer scales.
Weijters et al. (2010)	<i>International Journal of Research in Marketing</i>	A*	The effect of rating scale format on response styles: The number of response categories and response category labels	Scale-design elements (e.g., midpoint use and category labels) shape the mean, distribution, and reliability of self-reported responses.
Rocereto et al. (2011)	<i>Advances in International Marketing</i>	C	The role of response formats on extreme response style: A case of Likert-type vs. semantic differential scales	ERS is higher when respondents use Likert-type scales than when they use Semantic Differential scales, across United States and South Korean respondents. ERS is also significantly greater among US respondents compared to South Korean respondents when using Likert scales.
Harzing et al. (2012)	<i>Management International Review</i>	A	Response style differences in cross-national research: dispositional and situational determinants	Individuals with an East Asian (Chinese) cultural background are more likely to prefer mid-scale responses, whereas Western respondents (from Australia and Germany) are more likely to choose the scale's extreme options.
Wakita et al. (2012)	<i>Educational and Psychological Measurement</i>	A	Psychological distance between categories in the Likert scale: Comparing different numbers of options	The number of response options affected the perceived psychological distance between categories, especially for 7-point scales and items with socially undesirable content.
Li (2013)	<i>Expert Systems with Applications</i>	C	A novel Likert scale based on fuzzy set theory	The traditional Likert scale leads to information loss and distortion due to fixed response categories, while the proposed fuzzy Likert scale allows partial agreement through fuzzy membership values, capturing more accurate and nuanced responses.

(Continues)

Authors	Journal name	ABDC category	Title	Key findings
Weijters et al. (2013)	<i>Journal of Consumer Research</i>	A*	The effect of familiarity with the response category labels on item response to Likert scales	Respondents are more likely to select response categories that use labels they find familiar and common in day-to-day language.
Chen et al. (2015)	<i>Journal of Marketing Analytics</i>	C	What is the optimal number of response alternatives for rating scales? From an information processing perspective	The 5-point rating scale is the optimal format for questionnaire design because it provides the best balance between low cognitive effort, minimal response bias, and high information quality.
Eutsler and Lang (2015)	<i>Behavioral Research in Accounting</i>	A	Rating scales in accounting research: The impact of scale points and labels	Fully labelling all scale points reduces response bias, minimizes error and improves variance and statistical power. Results indicate that a 7-point scale maximizes variance more effectively than shorter or longer scales, and adding more than 7 points does not provide additional benefits.
Beuthner et al. (2018)	<i>International Journal of Market Research</i>	A	Examining survey response styles in cross-cultural marketing research: A comparison between Mexican and South Korean respondents	Mexican respondents were more likely to choose extreme responses, including strong disagreement, whereas South Korean respondents showed a stronger preference for midpoint responses.
Dolnicar (2021)	<i>Annals of Tourism Research</i>	A*	5/7-point “Likert scales” aren’t always the best option	Bipolar 5/7-point survey formats, often miscalled Likert scales, show low reliability, take longer to complete, and capture respondents’ personal or culture-specific response styles rather than true opinions.
Weijters et al. (2021)	<i>International Journal of Research in Marketing</i>	A*	Extremity in horizontal and vertical Likert scale format responses. Some evidence of how visual distance between response categories influences extreme responding	Data collected using scale formats with unequal spacing between response options (e.g., those commonly found in vertical vs. horizontal layouts) can result in variations in measurement model parameters, including residuals and factor loadings.

A2. Likert-type scale response anchors

Level of agreement		Level of appropriateness	
1.	Strongly disagree	1.	Absolutely inappropriate
2.	Disagree	2.	Inappropriate
3.	Somewhat disagree	3.	Slightly inappropriate
4.	Neither agree or disagree	4.	Neutral
5.	Somewhat agree	5.	Slightly appropriate
6.	Agree	6.	Appropriate
7.	Strongly agree	7.	Absolutely appropriate

(Continues)

Level of acceptability		Knowledge of action	
1.	Totally unacceptable	1.	Never true
2.	Unacceptable	2.	Rarely true
3.	Slightly unacceptable	3.	Sometimes but infrequently true
4.	Neutral	4.	Neutral
5.	Slightly acceptable	5.	Sometimes true
6.	Acceptable	6.	Usually true
7.	Perfectly acceptable	7.	Always true
Priority		Frequency	
1.	Not a priority	1.	Never
2.	Low priority	2.	Rarely, in less than 10% of the chances when I could have
3.	Somewhat priority	3.	Occasionally, in about 30% of the chances when I could have
4.	Neutral	4.	Sometimes, in about 50% of the chances when I could have
5.	Moderate priority	5.	Frequently, in about 70% of the chances when I could have
6.	High priority	6.	Very true of me
7.	Essential priority	7.	Every time
Level of satisfaction—7 points		Level of satisfaction—5 points	
1.	Completely dissatisfied	1.	Not at all satisfied
2.	Mostly dissatisfied	2.	Slightly satisfied
3.	Somewhat dissatisfied	3.	Moderately satisfied
4.	Neither satisfied or dissatisfied	4.	Very satisfied
5.	Somewhat satisfied	5.	Extremely satisfied
6.	Mostly satisfied		
7.	Completely satisfied		
Level of familiarity		Level of awareness	
1.	Not at all familiar	1.	Not at all aware
2.	Slightly familiar	2.	Slightly aware
3.	Somewhat familiar	3.	Somewhat aware
4.	Moderately familiar	4.	Moderately aware
5.	Extremely familiar	5.	Extremely aware
Level of difficulty		Likelihood	
1.	Very difficult	1.	Extremely unlikely
2.	Difficult	2.	Unlikely
3.	Neutral	3.	Neutral
4.	Easy	4.	Likely
5.	Very easy	5.	Extremely likely
Level of quality		Level of concern	
1.	Poor	1.	Not at all concerned
2.	Fair	2.	Slightly concerned
3.	Good	3.	Somewhat concerned
4.	Very good	4.	Moderately concerned
5.	Excellent	5.	Extremely concerned