SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 211

AN ITERATIVE METHOD FOR SURVEY IMPROVEMENT USING STATISTICAL ANALYSIS

Piotr GRZYBOWSKI[1], Dominika SIWIEC[2], Andrzej PACANA[3*]

[1] Rzeszow University of Technology, Faculty of Mechanical Engineering and Aeronautics; piotrg@prz.edu.pl, ORCID: 0000-0002-3207-5053
[2] Rzeszow University of Technology, Faculty of Mechanical Engineering and Aeronautics; d.siwiec@prz.edu.pl, ORCID: 0000-0002-6663-6621
[3] Rzeszow University of Technology, Faculty of Mechanical Engineering and Aeronautics; app@prz.edu.pl, ORCID: 0000-0003-1121-6352
* Correspondence author

Purpose: The purpose is to develop an iterative method for improving surveys (research questionnaire forms) using statistical analysis.
Design/methodology/approach: The method was developed based on a defined set of possible responses and their frequency, in order to identify questions that are incorrectly formulated or those for which the rating scale does not meet the assumed measurement criteria in subsequent stages. The process of survey analysis was automated, based on the average ratings, standard deviation, and properties of the normal distribution. Four main propositions were formed, methodically justified, and tested through an example.
Findings: Mapping customer requirements using the Likert scale may lead to data loss and bias in customer opinions in certain cases. Using a weighted, centred rating scale may cause excessive fragmentation of respondents' answers. Variance in the standard deviation of a sample analysed for a single point on the Likert scale leads to the loss of information.
Research limitations/implications: The method ensures the improvement of surveys before they are presented to a representative research group, increasing the chances of obtaining reliable and useful results. The proposed method can be successfully applied to analyse responses from any research questionnaires, especially in the early stages of their development.
Practical implications: The method helps identify potential errors during the development of preliminary versions of research questionnaires. Therefore, the method allows surveys to be adjusted before their use in obtaining a representative research sample. This increases the chances of obtaining reliable and useful results. This method can be applied by any survey designer in practice.
Social implications: The improved survey, according to this method, supports the process of obtaining respondent feedback in later stages of research, particularly during the main studies conducted among the target research sample. Respondents, when providing answers, have the opportunity to express their opinions more precisely, leading to the collection of accurate results.

Originality/value: The developed method, which is iterative in nature, dynamically identifies inconsistencies in respondents' answers and proposes alternative solutions for selecting rating scales or analysing respondents' answers.
Keywords: survey improvement, Likert scale, automated data analysis, rating scale design, production engineering.
Category of the paper: research paper.

1. Introduction

One of the key challenges in product improvement is capturing customer expectations, also known as the Voice of the Customer (VoC) (Siwiec et al., 2023). This stems from the general concept of the product design and improvement process, which involves identifying customer needs and taking further developmental actions based on them. This is traditionally carried out by gathering requirements, followed by their processing (analysis) and evaluation from, for example, a product quality perspective (Ma et al., 2017). Customer expectations are understood as their needs, opinions, and all requirements related to customer satisfaction with the usability of the product. At the same time, customer satisfaction is equated with product quality because it ranges from satisfaction to dissatisfaction with the product (Piasecka, Kłos, 2024; Sakao, 2007). Therefore, the higher the satisfaction, the higher the quality. The goal of businesses is to reduce the gap between satisfaction and dissatisfaction by adjusting product quality to meet both current and future market demands (Geng et al., 2021). To achieve this, it is initially essential to properly gather customer expectations in order to effectively guide further actions in designing new products or improving existing ones on the market (Szyjewski, 2018).

Customer preferences regarding a product focus on its attributes (features, criteria), often their importance to the customer and the level of satisfaction with their quality. Capturing and analysing these requirements can assist in determining the product that customers need. This approach also helps maintain a competitive position in the market (Qin et al., 2021; Siwiec et al., 2019). However, customer requirements are personalized and often diverse, especially in the case of mass-produced products. Another limitation is the varied terminology and semantics, which complicate mapping customer opinions during the design or improvement process. Additionally, new products often have new architectures, whose interpretation may be misinterpreted or imprecisely understood by customers. Ambiguous customer requirements are difficult to translate into design specifications (appropriate and understandable for designers) (Ma et al., 2017). Therefore, it is important that the initial stages of the design process, concerning the gathering of customer requirements, be carried out not only efficiently but also cost- and time-effective (Pacana, Siwiec, 2023; Wei et al., 2015). Defining what customers need is closely related to the need to use various techniques to support this process.

There are different techniques for gathering customer preferences. The choice of technique depends on the specific product being studied, including the research discipline. Requirement gathering techniques can be used individually or in an integrated manner, if necessary (Ma et al., 2017; Pacana, Siwiec, 2021). Some popular actions include, for example, market and customer research, including obtaining information about customers, consumer behaviour, or economic conditions (Huang et al., 2011). Therefore, customer requirements can be gathered directly from customers or through expert opinions from the respective industry (Geng et al., 2021; Xie et al., 2017). This is often done using surveys and interviews (Shan, Chen, 2011). Examples of applications of these techniques are presented in the literature, e.g. (Anuar, Mohd Yusuff, 2011; Jussani et al., 2018; Ostasz et al., 2022; Pacana et al., 2023; Schoenwitz et al., 2017; Sellitto et al., 2018). The Kano model is also popular, as it allows for the transformation of customer requirements into specialized language (Hwangbo et al., 2020; J. Li, Kim, 2023; X. Li et al., 2020; Neira-Rodado et al., 2020), or the technique of joint analysis (Qin et al., 2021), although these techniques require product evaluation by the customer. New emerging tools such as crowdsourcing platforms, which provide access to customer requirement gathering techniques and allow access to most customers worldwide, are also used (Ma et al., 2017).

Surveys, although widely used, are not without limitations. Even at the design stage, the knowledge and assumptions of the creators can influence the content, which may impact the construction of questions and suggest answers. For example, questions about colour preferences for product, limited to a closed set of options, may overlook colours that are truly desired by respondents, distorting the results.

Even a survey prepared with the utmost care—regarding semantic precision, scale suitability, or question complexity—may prove ineffective if the surveyed group lacks the necessary knowledge or competencies to correctly interpret the questions. In such cases, participants may provide incorrect answers, for example, choosing extreme or neutral options on the scale when they lack the possibility of providing a fully accurate response.

Conversely, incorrect survey design, such as improper question selection, inappropriate answer scales, or misalignment with the respondents' knowledge level, can lead to data that is not only useless but potentially harmful. The risk increases when the results of such surveys become the basis for strategic decisions, involving significant organizational resources.

This article presents an example of an iterative method for improving surveys, which can help identify potential errors at the stage of developing their initial versions. This allows for adjusting surveys before they are presented to a representative research group, thus increasing the chances of obtaining reliable and useful results.

2. Survey Research: Form, Scale, Advantages, and Limitations

As part of a synthetic literature review, it has been observed that the most popular techniques for capturing the Voice of the Customer (VoC) are survey studies (Anuar, Mohd Yusuff, 2011; Jussani et al., 2018; Ostasz et al., 2022; Schoenwitz et al., 2017). A significant tool supporting the survey process is the questionnaire, which facilitates data collection. Its preparation is characterized by strictly defined construction principles (Alwin, Krosick, 1991), including the method of data collection, such as qualitative methods, which involve answering questions like "how?" and "why?" or quantitative methods, which involve responding to questions like "how much?" and "how often?" (Krok, 2015). Ensuring the ability to obtain information from respondents is essential, while also tailoring the number and specificity of questions, maintaining clarity and precision in phrasing, and adjusting the response scale to the subject under study. This approach not only helps achieve reliable results but also maintains respondent engagement and prevents fatigue during the survey process (Szyjewski, 2018).

The first part of a survey typically serves as a brief introduction for respondents, explaining the purpose of the survey, the method of answering questions, the duration, anonymity of respondents, and access to anonymized survey results. The second part, known as the demographic section, helps to broadly identify who the respondent is, based on characteristics such as gender, age, residence, occupation, or interaction with the product (e.g., usage frequency) when the survey focuses on customer satisfaction with products. The third section contains the core questions developed to address the research questions posed (Krok, 2015; Szyjewski, 2018).

Proper formulation of research questions should follow these principles: ensuring grammatical, stylistic, and orthographic correctness; selecting appropriate vocabulary understandable to the target respondent group; minimizing the risk of misinterpretation; and avoiding leading questions. Survey responses are typically open-ended or closed-ended (Ponto, 2015). Open-ended questions allow respondents freedom of expression but should be constructed to elicit precise answers that do not pose challenges during result interpretation. Open-ended questions often take the form of brief (around two sentences) or extended responses (Szyjewski, 2018). Closed-ended questions, on the other hand, are so-called test questions. They present several possible answers, from which respondents can choose one (or more) according to their preferences. Closed-ended questions facilitate further analysis of survey data and contribute to increased objectivity (Memon et al., 2020). They also allow respondents not only to choose a preferred answer but also to determine its significance (Krok, 2015). This approach is used for alternative questions, where responses involve confirming or denying a given statement, such as "yes" or "no" or "true" or "false." A commonly used closed-ended question format is the Likert scale, often five- or seven-point (Konarski, Połomski, 2021).

The main advantages of surveys include the straightforward process of collecting customer expectations, low cost of obtaining a large volume of data, and ease of achieving satisfactory results. However, these advantages can also pose risks of obtaining inaccurate survey results, primarily due to human factors—the respondents participating in the survey. Additionally, errors during survey preparation, such as poorly formulated questions leading to misunderstandings or an inadequately designed response scale, may result in overly fragmented data (Krok, 2015; Szyjewski, 2018).

One common dilemma in designing survey questionnaires is the choice of a response scale. A poorly chosen scale can lead to excessive fragmentation or narrowing of survey results, impacting the reliability and credibility of measurements (Tarka, 2015; Wierzbiński et al., 2014).

Among the most popular rating scales in survey research is the Likert scale, frequently five- or seven-point (Cavaillé et al., 2024). Likert's original intent was to analyse latent phenomena and address limitations of simple scales by introducing multiple levels. This format allows assessment of not just a single test item but a series of aggregated items forming a construct. Consequently, the Likert scale ensures indirect measurement. It is particularly useful in studies examining attitudes, capturing their direction and strength, such as internal human thought processes, which are complex constructs not suitable for single-level scales. Furthermore, indirect measurement using a simple scale (single response) would yield less reliable results in such cases (Tarka, 2015; Weijters et al., 2021; Wierzbiński et al., 2014).

The Likert scale enables researchers to detect subtle differences, and its results do not significantly affect the final measurement outcome. This means that the scale items balance each other, unlike simple scales, where a single response can heavily influence the outcome (Tarka, 2015).

Common types of Likert scales include five- or seven-point scales (Dolnicar, 2021). Scales with two, three, or even nine levels are also possible. Depending on the needs, an appropriate scale can be selected; however, an overly extensive range may negatively impact the precision of respondent answers (Solis et al., 2022). Additionally, the greater the number of response options, the harder it is for typical respondents to discern differences between them, potentially leading to "response flattening", where scale points are merged and interpreted collectively (Capuano et al., 2016; Tarka, 2015).

3. Method

3.1. Determining the Set of Responses and Its Cardinality

A survey containing only closed-ended questions, where the possible responses are expressed using a scale (e.g., Likert scale), will have a finite number of possible answers. This allows for defining a finite set of all possible response combinations, the cardinality of which, denoted as **card(A)**, indicates the total number of possible responses.

The cardinality is significant for interpreting results. Too small a cardinality might omit crucial information, while too large a cardinality complicates data analysis, even with advanced computational tools. For example, let's consider survey questions about the safety of shipping methods:

1. Is air freight safe? NO/ YES
2. Is rail freight safe? NO/ YES

We can define the set of responses for question 1 (x_1) and question 2 (x_2) (1):

$$x_1 = \{NO, YES\}$$
$$x_2 = \{NO, YES\} \tag{1}$$

Thus, the set of all possible responses ***A*** consists of pairs of answers: (2):

$$A = \{\{NO, NO\}, \{NO, YES\}, \{YES, NO\}, \{YES, YES\}\} \tag{2}$$

The number of possible responses in this case is (3):

$$card(A) = card(x_1) \cdot card(x_2) = 2 \cdot 2 = 4 \tag{3}$$

This is an example of a survey where the questions are independent, meaning one cannot directly interpret responses to infer preferences (e.g., assuming customers prefer rail over air freight). However, analysis of results can reveal insights, such as "most respondents do not prefer rail freight while a majority prefer air freight". It is also possible to compute a correlation coefficient to determine whether a dependency exists between the responses.

Using the Likert scale expands the response set for each question to 5 options. Reformulating the example:

1. Is air freight safe? <1,5>,
   where 1 = STRONGLY DISAGREE, 5 = STRONGLY AGREE
2. Is rail freight safe? <1,5>,
   where 1 = STRONGLY DISAGREE, 5 = STRONGLY AGREE

The response sets for question 1 (x_1) and question 2 (x_2) become (4):

$$x_1 = \{1,2,3,4,5\}$$
$$x_2 = \{1,2,3,4,5\} \tag{4}$$

The number of possible responses is (5):

$$card(A) = card(x_1) \cdot card(x_2) = 5 \cdot 5 = 25 \tag{5}$$

For the given example a set of all possible elements (potential answers) of set ***A*** has been shown in the table 1.

Table 1.
A set A elements with corresponding possible answers

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{16}
x_1	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4
x_2	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1

	A_{17}	A_{18}	A_{19}	A_{20}	A_{21}	A_{22}	A_{23}	A_{24}	A_{25}
x_1	4	4	4	4	5	5	5	5	5
x_2	2	3	4	5	1	2	3	4	5

The greater the number of independent questions and the more granular the scale, the larger the cardinality of ***A***. For example, a survey with 80 Likert-scale questions yields (6):

$$card(A) = \prod_{i=1}^{80} card(x_i) = card(x_1) \cdot card(x_2) \cdot \cdots \cdot card(x_{80}) = 5^{80} \approx 8.271e + 55 \tag{6}$$

Replacing 10 questions with a binary scale reduces the cardinality to (7):

$$card(A) = \prod_{i=1}^{80-10} card(x_i) \cdot \prod_{i=1}^{10} card(y_i) = 5^{70} \cdot 2^{10} \approx 8.67361e + 51 \tag{7}$$

Using a binary scale for all questions gives (8):

$$card(A) = \prod_{i=1}^{80} card(y_i) = 2^{80} \approx 1.2089e + 24 \tag{8}$$

Reducing the number of Likert-scale questions to 30 results in (9):

$$card(A) = \prod_{i=1}^{30} card(x_i) = 5^{30} \approx 9.31322574615e + 20 \tag{9}$$

Cardinality can be significantly reduced by carefully choosing scales and limiting the number of questions (e.g., reducing 80 to 30). This optimization is essential for computational efficiency in analysing correlations between survey responses. Reducing the number of questions also shortens the time required to complete the survey, saving respondents' time.

The main challenge is deciding which questions to retain or modify and how to adjust scales. These decisions cannot be made solely during the survey's design phase. A critical step is conducting preliminary validation, such as testing the survey on a smaller group of respondents and analysing their responses through pilot studies.

3.2. Automated survey analysis – initial assumptions

Automating the survey analysis process is crucial when dealing with a large number of questions. This article adopts the following assumptions for refining surveys:

1. Remove questions that do not provide any non-obvious insights to the surveyor based on the analysis of results.
2. Remove questions that respondents do not understand.
3. Remove questions containing propositions that, regardless of the respondent's answer, can be used to draw conclusions (e.g., the question: "Will you stop using biohazard materials in the production process?" with possible answers YES/NO allows the surveyor to infer that 100% of respondents previously used biohazard materials in the production process, regardless of their answers).

Additionally, to ensure the analysis process remains accurate, the following types of completed surveys should be excluded from the dataset:

1. Surveys filled out carelessly, such as marking the maximum value on the scale for every question.

Responses that deviate significantly from the majority but demonstrate internal consistency in justification (e.g., due to a respondent's extraordinary insight) should be examined in detail.

It is evident that performing such an analysis requires defining a function (10):

$$f: X \to Y \tag{10}$$

This function maps survey responses to numerical values, for example, for the set (11):

$$B = \{\{NO,NO\},\{NO,YES\},\{YES,NO\},\{YES,YES\}\} \tag{11}$$

Table 2 presents an example of such mapping.

Table 2.
Example of assigning numerical values to survey responses in binary scale

x	f(x)
NO	0
YES	1

Transforming the data using this function produces the set ***C*** with the following possible responses (12):

$$C = \{\{0,0\},\{0, 1\}, \{1, 0\}, \{1,1\}\} \tag{12}$$

During such transformations, it is crucial to avoid distortions that could affect further analysis. Table 3 demonstrates an example of two functions, ***f(x)*** and ***g(x)***, where the calculated mean values differ significantly despite identical input values ***x.***

Table 3.
Example of assigning numerical values to survey responses using functions

x	f(x)	g(x)
Strongly disagree	1	0
Disagree	2	1
Neutral	3	2
Agree	4	3
Strongly agree	5	5

Assuming the set of responses (13):

$$E = \{\text{Strongly disagree, Disagree, Neutral, Agree, Strongly agree}\} \tag{13}$$

Transforming the set using ***f*** yields (14):

$$F = \{1,2,3,4,5\} \tag{14}$$

The mean value for this transformed set is (15):

$$S_F = 3 \tag{15}$$

Transforming the set using ***g*** yields (16):

$$G = \{0,1,2,3,5\} \tag{16}$$

The mean value for this set is (17):

$$S_G = 2.2 \tag{17}$$

For subsequent considerations, it is assumed that the transformation function ***h***, which maps response sets to numerical values, is linear. It is also assumed that the response set inherently reflects the respondent's agreement level with the survey proposition in a linear fashion and excludes responses that could skew the mean value. An example of an improperly constructed scale would be the set (18):

$$E = \{\text{Neutral, Rather agree, Agree, Highly Agree, Strongly agree}\} \tag{18}$$

3.3. Automated Analysis – Survey Assessment Tools

It is proposed that, during analysis, the initial assumption should be that survey questions are independent (no correlation between responses) and linear. In this case, the **mean** of respondents' answers is expressed by formula (19):

$$S_i = \frac{1}{N}\sum_{j=1}^{N} x_j \tag{19}$$

where:
S_i – the mean value for responses to the ***i-th*** question in the survey,
N – the total number of responses to the ***i-th*** question in the survey,
x_j – the value of the response to the ***i-th*** question by the ***j-th*** respondent.

This formula represents one of the simple, widely-used mathematical tools to evaluate the central tendency.

Another mathematical tool applied in further analysis is the **standard deviation from the mean**, expressed by formula (20):

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{N} x_j^2}{N} - \mu_i} \tag{20}$$

where:
σ – the standard deviation from the mean,
N – the total number of responses to the ***i-th*** question in the survey,
x_j – the value of the response to the ***i-th*** question by the ***j-th*** respondent,
μ_i – the expected value of responses to the ***i-th*** question in the survey.

When creating survey questions along with the available answer options for respondents, the survey creator defines the set of possible responses and, consequently, the scale midpoint. Assuming the distribution of responses can be approximated using a normal distribution, the assumption can be made that (21):

$$S_i = \mu_i \tag{21}$$

where:
μ_i – the expected value of responses to the ***i-th*** question in the survey,
S_i – the mean value for responses to the ***i-th*** question in the survey.

At the same time, the calculated mean S_i based on responses from test surveys, may deviate from the expected value μ_i This leads to the formulation of the following postulate:

Postulate 1: Analyse questions for which the mean value of survey responses deviates from the expected value assumed by the survey creator.

Justification: For a normal distribution, 99.7% of responses should fall within a range of $\mathbf{3 \cdot \sigma}$ from the expected value (the so-called three-sigma rule). If the mean value is not at the midpoint of the scale (i.e., it differs from the expected value μ), this could indicate that the function $\boldsymbol{h}$ is not properly defined, the set of responses does not fully reflect respondents' preferences (as it might not allow responses beyond the predefined scale), or responses cannot be described using a normal distribution.

Another important issue to address, considering the properties of a normal distribution, arises when the calculated value of $\mathbf{3 \cdot \sigma}$ indicates that the assumed maximum of the response set ***U*** (22):

$$\boldsymbol{max(U) \ll S_i + 3 \cdot \sigma} \tag{22}$$

Postulate 2: Analyse questions for which $\boldsymbol{S_i + 3 \cdot \sigma_i}$ significantly exceeds the maximum scale value ***max(U)***.

Justification: Similar to the previous point, for a normal distribution, 99.7% of responses should fall within $\boldsymbol{3 \cdot \sigma_i}$ of the expected value. If this value exceeds the maximum scale, it suggests that responses exceeding the predefined scale could be overlooked during the actual survey.

Conversely, there is also the possibility that the scale used is excessively granular. Such a situation would be indicated by the standard deviation (23):

$$\boldsymbol{max(U) \gg S_i + 3 \cdot \sigma} \tag{23}$$

Postulate 3: Analyse questions for which the maximum scale value ***max(U)*** is significantly greater than $\boldsymbol{S_i + 3 \cdot \sigma}$ This indicates that the majority of respondents do not utilize the full scale, necessitating its adjustment.

Justification: Failure to fully utilise the scale suggests that the scale should be modified. It may be possible to reduce unused scale elements, decreasing the set of possible response combinations for the survey questions.

Considering that both mean values and standard deviations are influenced by respondents who may not complete the survey diligently, such instances of unreliability can be detected by analysing response results for individual questions in the context of the mean value and standard deviation.

Postulate 4: Since 95% of responses should fall within $\boldsymbol{S_i \pm 2 \cdot \sigma_i}$ define a transformation function $\boldsymbol{x_j \to v_j}$ such that (24):

$$v_j = \begin{cases} 0, if\ \boldsymbol{S_i - 2 \cdot \sigma_i < x_j\ < S_i + 2 \cdot \sigma_i} \\ 1, if\ \boldsymbol{S_i - 2 \cdot \sigma_i > x_j\ \cup\ S_i - 2 \cdot \sigma_i < x_j} \end{cases} \tag{24}$$

Then for the ***j-th*** respondent, given their set of responses $\boldsymbol{X_j}$, create the set of transformed elements $\boldsymbol{V_j}$. If the majority of elements in this set equal 1, it indicates that the respondent's answers do not align with the central tendency for most questions. Consequently, their survey responses should be reviewed in detail.

Justification: Respondents whose answers deviate significantly from the central tendency might either have deep insights and form unconventional conclusions based on logical reasoning, or they might have completed the survey carelessly (e.g., always selecting the maximum or minimum response).

4. Results

The developed iterative survey improvement method was tested using preliminary research. A questionnaire with a five-point Likert scale was used to survey 25 customers. The survey included two main questions, resulting in a total of 80 possible single-choice responses. Due to the illustrative nature of the study and method testing, the results are presented in an anonymised form.

Using formulas (19-21), the mean of the respondents' answers and the standard deviation from the mean were calculated. It was subsequently assumed that the distribution of responses could be approximated using a normal distribution. Following Postulate 1, questions for which the mean value of the survey results deviated from the expected value assumed by the survey creator were analysed. Then, according to formula (22), the maximum of the response set was estimated. The results are presented in Table 4.

Table 4.
Fragment of the results of an anonymised research survey with a Likert scale

No.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	4	3	2	2	3	3	2	2	4	2
2	4	3	3	3	5	3	2	3	4	1
3	4	2	3	3	5	4	2	3	5	2
4	3	2	3	4	4	3	2	2	5	2
5	5	5	5	5	5	4	2	1	2	2
6	3	3	3	3	3	3	3	3	3	3
7	4	2	2	3	5	4	3	2	4	2
8	4	2	3	4	3	4	5	2	5	3
9	2	2	2	2	4	3	1	2	3	2
10	5	4	4	5	5	3	2	2	5	2
11	2	1	2	2	4	3	2	1	5	1
12	5	5	5	5	2	2	2	2	5	2
13	2	2	2	2	2	1	4	2	2	2
14	4	4	4	4	2	1	1	1	2	1
15	2	2	2	2	2	2	2	2	2	2
16	2	1	1	2	3	3	3	1	2	2
17	4	2	3	2	3	2	1	2	5	2
18	4	2	3	4	5	4	2	1	5	2
19	4	3	3	2	5	4	3	2	3	2
20	4	2	2	3	5	3	2	3	5	1
21	3	2	4	3	3	3	2	2	4	3
22	4	2	2	3	3	2	1	2	3	2
23	5	1	1	3	4	3	2	2	4	2
24	3	3	2	4	4	4	3	3	3	3
25	4	3	2	2	4	3	2	2	5	1
S	3.60	2.52	2.72	3.08	3.72	2.96	2.24	2.00	3.80	1.96
σ	1.00	1.08	1.06	1.04	1.10	0.89	0.93	0.65	1.19	0.61
3σ	6.60	5.77	5.90	6.19	7.02	5.63	5.02	2.00	3.80	1.96

where: Q1-Q10 – responses to the survey questions, **S** – the mean value for the response to the ***i-th*** question in the survey, ***σ*** – the standard deviation from the mean, 3σ – three sigma value.

Analysing the values of question Q5 (highlighted in red), it can be noted that the design does not match the mean value of 3.72. A deeper analysis of the responses suggests that none of the respondents chose answer 1, while there are an equally large number of responses for answers 4 and 5. This raises the question of whether the scale was set properly and if it should be modified. As shown for questions Q8 and Q10 (highlighted in yellow), respondents did not use the full scale designed to answer the questions; therefore, the scale could potentially be reduced, limiting the set of possible answers.

Considering that the mean values and standard deviations are influenced by responses from respondents who may not have filled out the survey in a reliable manner, an attempt was made to detect such cases of unreliability by analysing the answers to individual questions in the context of the mean and standard deviation values. Formula (24) was used for this purpose. The result, which checks whether the results fall within 2σ of the standard deviation (0 = yes, 1 = no), is presented in Table 5.

Table 5.
The result checking the survey results according to postulate 4

No.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	1	1	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	0	0	0
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0
12	0	1	1	0	0	0	0	0	0	0
13	0	0	0	0	0	1	0	0	0	0
14	0	0	0	0	0	1	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0
Sum of deviations	0	2	2	0	0	2	1	0	0	0
Acceptable number of responses	2	2	2	2	2	2	2	2	2	2
Max. rating in the survey	5	5	5	5	5	5	5	5	5	5
Min. rating in the survey	1	1	1	1	1	1	1	1	1	1
Automatically determined max. ratings in the survey	5	5	5	5	6	4	3	3	6	3
Automatically determined min. ratings in the survey	3	1	1	1	2	2	1	1	2	1
Is the scale appropriate?	Analyse			OK	Analyse					
Proposed max.	5	5	5	5	5	4	5	3	5	3
Proposed min.	2	1	1	2	2	1	1	1	2	1

where: Q1-Q10 – responses to the survey questions.

As assumed, if the majority of elements in the analysed set of responses were equal to 1, it was considered that the respondent did not provide answers consistent with the central tendency for most of the questions. Therefore, a detailed analysis of the survey results for such a respondent was recommended. These respondents may be providing answers based on a deep understanding of the issue and formulating non-standard conclusions based on their reasoning process, or they may indicate respondents who completed the survey carelessly (e.g., always selecting the maximum or minimum answer). Such a phenomenon was observed in the example answers (responses to questions: Q1-Q3, Q5-Q10). Then, according to the automated analysis, the proposed maximum and minimum scores should be considered for these types of responses. Based on the degrees of freedom derived from the new range of assigned scores (max. and min.), the adequacy of the rating scale is checked for five-point, four-point, and three-point scales. Even with the presence of deviations, the potential of the previously adopted rating scale is observed without the need for its change/narrowing, e.g., for Q2, Q3, Q7. In other cases, it is proposed to change the scale to a four-point or three-point scale. This confirms the effectiveness of the proposed method.

5. Discussion and Conclusion

The process of gathering customer expectations through survey questionnaires, while widely regarded as the most popular method, still faces certain limitations (Krok, 2015; Szyjewski, 2018). Challenges arise not only during the design phase of the survey but also after its completion—during the interpretation of results. The survey design and the reliability of responses provided by respondents are closely interrelated. Even with the utmost care in designing the questionnaire, such as systematic question selection or response options, inadequate answers may result from respondents' lack of knowledge, competence, or low engagement during the survey process (Tarka, 2015; Weijters et al., 2021; Wierzbiński et al., 2014). This leads to unreliable data, creating challenges for subsequent decision-making processes based on this data.

To mitigate these limitations from the initial stages of survey development, an iterative method was designed to improve questionnaires using statistical analysis. This method helps identify potential errors during the preparation of preliminary survey versions.

The method was tested using a pilot survey that included two main questions, with 80 possible responses. The survey was conducted on a preliminary sample of 25 respondents.

Analysing the results of the proposed method, including comparisons of its efficiency with other studies (e.g., Dolnicar, 2021; Solis et al., 2022; Weijters et al., 2021; Westland, 2022), yielded the following conclusions and observations:

1. Mapping customer requirements using a Likert scale may result in data loss and biased customer opinions in certain cases.
2. Using a weighted, centred rating scale, such as a five-point scale, does not always provide reliable results, occasionally leading to excessive fragmentation of respondents' answers.

3. The occurrence of variance in the standard deviation of a sample analysed for a single point on the Likert scale generates information loss.
4. When respondents' opinions are polarized, and they are uncertain about their stance (e.g., strongly agreeing or strongly disagreeing with a given statement), they tend to choose the midpoint of the Likert scale.
5. Strongly held beliefs, whether positive or negative, that deviate significantly from other data may indicate either hasty (or inaccurate) responses from respondents or the presence of an ideal case, often referred to as the "genius" solution dedicated to the studied phenomenon.

It has been suggested that in preliminary surveys, the degree of question comprehension can be assessed by asking an additional question, such as: "Is the question clear to you, and do the answers allow you to express your opinion on the topic?" However, there is concern that respondents may prioritize quickly completing the survey, leading to a careless approach where most questions and/or answers are deemed unclear.

The developed iterative method has been shown to support questionnaire refinement before presenting it to a representative research group, increasing the likelihood of obtaining reliable and useful results. The proposed questionnaire can be successfully applied to analyse responses from various research surveys, particularly during their early development stages.

References

1. Alwin, D.F., Krosick, J.A. (1991). The Reliability of Survey Attitude Measurement. *Sociological Methods & Research*, *20*(1), 139-181. https://doi.org/10.1177/004912419102000100 5
2. Anuar, A., Mohd Yusuff, R. (2011). Manufacturing best practices in Malaysian small and medium enterprises (SMEs). *Benchmarking: An International Journal*, *18*(3), 324-341. https://doi.org/10.1108/14635771111137750
3. Capuano, A.W., Dawson, J.D., Ramirez, M.R., Wilson, R.S., Barnes, L.L., Field, R.W. (2016). Modeling Likert Scale Outcomes With Trend-Proportional Odds With and Without Cluster Data. *Methodology*, *12*(2), 33-43. https://doi.org/10.1027/1614-2241/a000106
4. Cavaillé, C., Chen, D.L., Van der Straeten, K. (2024). Who cares? Measuring differences in preference intensity. *Political Science Research and Methods*, 1-17. https://doi.org/10.1017/psrm.2024.27
5. 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

6. Geng, L., Shi, X., Zu, L., Chai, M., Xing, J. (2021). Importance Calculation of Customer Requirements for Incremental Product Innovation. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.633472
7. Huang, H.-Z., Li, Y., Liu, W., Liu, Y., Wang, Z. (2011). Evaluation and decision of products conceptual design schemes based on customer requirements. *Journal of Mechanical Science and Technology*, *25*(9), 2413-2425. https://doi.org/10.1007/s12206-011-0525-6
8. Hwangbo, Y., Yang, Y.-S., Kim, M.-S., Kim, Y. (2020). The Effectiveness of Kano-QFD Approach to Enhance Competitiveness of Technology-Based SMEs through Transfer Intention Model. *Sustainability*, *12*(19), 7885. https://doi.org/10.3390/su12197885
9. Jussani, A.C., Vasconcellos, E.P.G. de, Wright, J.T.C., Grisi, C.C. de H.E. (2018). Marketing internationalization: influence factors on product customization decision. *RAUSP Management Journal*, *53*(4), 555-574. https://doi.org/10.1108/RAUSP-07-2018-0043
10. Konarski, R., Połomski, P. (2021). Specyfika badań ankietowych. Rodzaje pytań, opracowanie ankiet, organizacja badań, analiza wyników, trafność, rzetelność, tworzenie skali. In: *Epidemiologia. Od teorii do praktyki* (pp. 114-126). PZWL.
11. Krok, E. (2015). Survey questionnaire structure and its impact on the research results. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Studia Informatica*, *37*, 55-73. https://doi.org/10.18276/si.2015.37-05
12. Li, J., Kim, K. (2023). Kano-QFD-based analysis of the influence of user experience on the design of handicraft intangible cultural heritage apps. *Heritage Science*, *11*(1), 59. https://doi.org/10.1186/s40494-023-00903-w
13. Li, X., Zhu, S., Yüksel, S., Dinçer, H., Ubay, G.G. (2020). Kano-based mapping of innovation strategies for renewable energy alternatives using hybrid interval type-2 fuzzy decision-making approach. *Energy*, *211*, 118679. https://doi.org/10.1016/j.energy.2020.118679
14. Ma, X.-J., Ding, G.-F., Qin, S.-F., Li, R., Yan, K.-Y., Xiao, S.-N., Yang, G.-W. (2017). Transforming Multidisciplinary Customer Requirements to Product Design Specifications. *Chinese Journal of Mechanical Engineering*, *30*(5), 1069-1080. https://doi.org/10.1007/s10033-017-0181-6
15. Memon, M.A., Ting, H., Cheah, J.-H., Thurasamy, R., Chuah, F., Cham, T.H. (2020). Sample Size for Survey Research: Review and Recommendations. *Journal of Applied Structural Equation Modeling*, *4*(2), i-xx. https://doi.org/10.47263/JASEM.4(2)01
16. Neira-Rodado, D., Ortíz-Barrios, M., De la Hoz-Escorcia, S., Paggetti, C., Noffrini, L., Fratea, N. (2020). Smart Product Design Process through the Implementation of a Fuzzy Kano-AHP-DEMATEL-QFD Approach. *Applied Sciences*, *10*(5), 1792. https://doi.org/10.3390/app10051792

17. Ostasz, G., Siwiec, D., Pacana, A. (2022). Model to Determine the Best Modifications of Products with Consideration Customers' Expectations. *Energies*, *15*(21), 8102. https://doi.org/10.3390/en15218102
18. Pacana, A., Siwiec, D. (2021). Analysis of the Possibility of Used of the Quality Management Techniques with Non-Destructive Testing. *Tehnicki Vjesnik - Technical Gazette*, *28*(1). https://doi.org/10.17559/TV-20190714075651
19. Pacana, A., Siwiec, D. (2023). Method of Fuzzy Analysis of Qualitative-Environmental Threat in Improving Products and Processes (Fuzzy QE-FMEA). *Materials*, *16*(4), 1651. https://doi.org/10.3390/ma16041651
20. Pacana, A., Siwiec, D., Bednárová, L., & Petrovský, J. (2023). Improving the Process of Product Design in a Phase of Life Cycle Assessment (LCA). *Processes*, *11*(9), 2579. https://doi.org/10.3390/pr11092579
21. Piasecka, I., Kłos, Z. (2024). An Assessment of the Environmental Impact of Construction Materials of Monocrystalline and Perovskite Photovoltaic Power Plants Toward Their Sustainable Development. *Materials*, *17*(23), 5787. https://doi.org/10.3390/ma17235787
22. Ponto, J. (2015). Understanding and Evaluating Survey Research. *J Adv Pract Oncol*, *6*(2), 168–171.
23. Qin, Y., Zhaofa, Y., Xuzheng, L., Zufang, Z., Weijie, C., Sheng, R. (2021). Product Module Attribute Parameter Configuration Model considering Customer Requirements Preferences. *Mathematical Problems in Engineering*, 1-10. https://doi.org/10.1155/2021/6632057
24. Sakao, T. (2007). A QFD-centred design methodology for environmentally conscious product design. *International Journal of Production Research*, *45*(18-19), 4143–4162. https://doi.org/10.1080/00207540701450179
25. Schoenwitz, M., Potter, A., Gosling, J., Naim, M. (2017). Product, process and customer preference alignment in prefabricated house building. *International Journal of Production Economics*, *183*, 79-90. https://doi.org/10.1016/j.ijpe.2016.10.015
26. Sellitto, M.A., Nunes, F.L., Vallladares, D.R.F. (2018). Factors that contribute to the use of modularisation in the automotive industry: a survey in brazil. *South African Journal of Industrial Engineering*, *29*(4). https://doi.org/10.7166/29-4-1946
27. Shan, Q., Chen, Y. (2011). Product Module Identification Based on Assured Customer Requirements. *Procedia Engineering*, *15*, 5313-5317. https://doi.org/10.1016/j.proeng.2011.08.985
28. Siwiec, D., Bednárová, L., Pacana, A., Zawada, M., Rusko, M. (2019). Decision support in the selection of fluorescent penetrants for industrial non-destructive testing. *Przemysł Chemiczny*, *1*(10), 92-94. https://doi.org/10.15199/62.2019.10.12
29. Siwiec, D., Pacana, A., Gazda, A. (2023). A New QFD-CE Method for Considering the Concept of Sustainable Development and Circular Economy. *Energies*, *16*(5), 2474. https://doi.org/10.3390/en16052474

30. Solis, D., Hutchinson, D., Longnecker, N. (2022). Visual Discrete Format: An Alternative to Likert-Type Formats of Survey Items Sensitive Enough to Measure Small Changes in Stable Constructs Such as Self-Concept in Science. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)*, *10*(2), 1-16. https://doi.org/10.23947/2334-8496-2022-10-2-01-16
31. Szyjewski, Z. (2018). Methods of studying social behavior. *Studia i Prace WNEiZ*, *54*, 147-162. https://doi.org/10.18276/sip.2018.54/2-10
32. Tarka, P. (2015). Własności 5- i 7-stopniowej skali Likerta w kontekście normalizacji zmiennych metodą Kaufmana i Rousseeuwa. *Prace Naukowe Uniwersytetu Ekonomicznego We Wrocławiu*, *385*. https://doi.org/10.15611/pn.2015.385.31
33. Wei, Wei., Liu, A., Lu, S.C.-Y., Wuest, T. (2015). Product Requirement Modeling and Optimization Method Based on Product Configuration Design. *Procedia CIRP*, *36*, 1-5. https://doi.org/10.1016/j.procir.2015.01.020
34. Weijters, B., Millet, K., Cabooter, E. (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*(1), 85-103. https://doi.org/10.1016/j.ijresmar.2020.04.002
35. Westland, J.C. (2022). Information loss and bias in likert survey responses. *PLOS ONE*, *17*(7), e0271949. https://doi.org/10.1371/journal.pone.0271949
36. Wierzbiński, J., Kuźmińska, A.O., Król, G. (2014). Consequences of using different types of rating scales. *Problemy Zarzadzania*, *12*(1), 113-136. https://doi.org/10.7172/1644-9584.45.8
37. Xie, M.-Q., Jiang, Q.-Q., Cheng, W.-P., Ma, X.-X. (2017). Determining the Importance Ratings of Customer Requirements of Automotive Clutch Based on Quality Function Deployment. *DEStech Transactions on Engineering and Technology Research*, *mdm*. https://doi.org/10.12783/dtetr/mdm2016/4932