JOURNAL
of DESIGN
THINKING

Comparison Between Discrete and Analog Semantic Differential Scales Accuracies in Kansei Engineering (Case Study: Reception Chairs)

***Ali Dasmeh*[1], *Nasser Koleini Mamaghani*[2*], *Peyman Hassani-Abharian*[3]**

[1] Human Development Studies, Max Planck Institute, Berlin, Germany. Email: dasmeh@mpib-berlin.mpg.de
[2] Industrial Design Department, School of Architecture and Environmental Design, Iran University of Science and Technology (IUST), Tehran, Iran. Email: koleini@iust.ac.ir
[3] Department of Cognitive Psychology and Rehabilitation, Institute for Cognitive Science Studies, Tehran, Iran. Email: abharian@icss.ac.ir

[*]**Corresponding author:** Nasser Koleini Mamaghani

DOI: 10.22059/JDT.2025.388565.1137

Received: *10 January 2025,* ***Revised:*** *29 January 2025,* ***Accepted:*** *29 January 2025,* ***Available Online*** *from 29 January 2025.*

Abstract

This study aims to compare the distribution of semantic differential scale results, both discrete and analog, used in Kansei Engineering studies. Previous research shows that observers have a tendency to choose the extreme ends of Likert or discrete semantic differential rating scales. Conventionally, the discrete semantic differential rating scale is used in Kansei Engineering and the above-mentioned tendency may affect the result of studies. Non-normality of the distribution can be an indicator of the presence of error and bias in that method and mitigate its validity. The data was collected through a Kansei Engineering process using a real-world case study of reception chairs. The distribution of the results for both rating scales can be considered normal, hence, the previous researches' achievements are not valid in the field of Kansei Engineering. This may be due to the ability to compare each answer with others, even implicitly, which is possible in the Kansei form. The results show that the difference between the average scores of the two rating scales is significant (t-test, $p < 0.05$), warranting further investigation.

Keywords

Likert Scale, Kansei Engineering, Rating Scale, Normal Distribution, Reception Chair.

Introduction

Designing based on human emotion is crucial for creating products that resonate deeply with consumers. Emotions play a significant role in our decision-making processes, often influencing our choices more than purely rational factors. By understanding and incorporating emotional responses into design, we can create products that not only meet functional needs but also evoke positive feelings, enhance user experiences, and foster brand loyalty. This approach recognizes that successful design is not just about aesthetics or practicality; it's about creating meaningful connections with users on an emotional level (Nagamachi, 1995).

As various sections of industries realized the power of emotional connections in driving consumer behavior and building brand loyalty, they began to adopt Kansei/Affective Engineering to create products that stood out in the competitive marketplace. This approach has become increasingly important in industries like automotive, consumer electronics, and interior design, where emotional appeal can significantly influence consumer preferences (Nagamachi, 1999).

Kansei Engineering was developed in Japan to translate consumers' feelings about the product into specific design decisions (Nagamachi, 1995). KE utilizes statistical techniques to analyze customer feedback and identify key relationships between subjective preferences and objective product attributes. Kansei Engineering benefits from a modified version of the semantic differential method in data collection processes (Nagamachi, 2010). [Semantic differential] Five points (five bipolar pairs of adjectives) have been proven to yield reliable findings, which highly correlate with alternative Likert numerical measures of the same attitude (Osgood, 1957). Developing self-report Likert scales is essential to modern psychology (Jebb et al., 2021). Likert scaling, introduced by Rensis Likert, is the most widely used method of measuring personality, social, and psychological attitudes. The popularity of Likert scales can be traced to several factors, including ease of construction, intuitive appeal, adaptability, and usually good reliability (Hodge & Gillespie, 2003).

Likert response surveys are widely applied in marketing, public opinion polls, epidemiological, and economic disciplines. Theoretically, Likert mapping from real-world beliefs could lose significant amounts of information, as they are discrete categorical metrics. Arguments and counterexamples are provided to show how this loss and bias can potentially be substantial under extreme polarization or strong beliefs held by the surveyed population, and where the survey instruments are poorly controlled. The discussion and conclusions argue that further revisions to survey protocols can ensure that information loss and bias in Likert-scaled data are minimal (Westland, 2022).

One important problem in the measurement of non-cognitive characteristics such as personality traits and attitudes is that it has traditionally been made through Likert scales, which are susceptible to response biases such as social desirability (SDR) and acquiescent (ACQ) responding. Response-style-induced errors of measurement can affect the reliability estimates and overestimate convergent validity by correlating higher with other Likert-scale-based measures (Kreitchmann et al., 2019).

The study of chairs as a case study in Kansei Engineering (KE) has a rich historical context, both in conventional and novel methodologies. Hsu et al. (2017) explored the induction of Kansei (affect) from 3D-rendered chair forms, employing eye tracking techniques as an alternative to the traditional Kansei Engineering procedure. Notably, this research team had previously utilized semantic differential scales in their investigations with the same case study. Zhou et al. (2023) presented a Kansei engineering study on electric recliner chairs, introducing a novel computational method.

While their data collection approach remained conventional, their computational work aimed to address the limitations of the semantic differential method. This research was conducted on reception chairs because its samples were diverse on online shops enabling researchers to select more real and suitable images of products for every permutation of items and categories in the KE.

This research aims to address one of the most critical accuracy aspects of the Kansei Engineering process which is collecting consumers' emotions when dealing with products.

Kansei Engineering could be used to identify which design characteristics influence users' perception of whether or not a design is modern (Tama et al., 2015). The word *modern* is an instance of an adjective, any adjective can be replaced. Kansei decomposition is used to link the desired perceptions of a design with senses and engineering characteristics, so this decomposition identifies which engineering decisions may affect user perception of a design (Ayas, 2011). In the terminology of Kansei engineering, an item implies the design item of the sample product, and a category means the detail of the design item. For instance, color, shape, size, roundness, and so forth are examples of items; and red, yellow, green, blue, and so forth are the categories for the color item. The Kansei/affective engineer should be very careful of the sample product's items and categories (Nagamachi, 2010).

The accuracy and validity of a rating scale have been among the most important concerns of all researchers, so they proposed different methods to address this concern.

Kusmaryono et al. (2022) through 60 published surveys by a systematic literature review concluded the use of a rating scale with an odd number of responses of more than five points (especially on a seven-point scale) is the most effective in terms of reliability and validity coefficients. Based on their output they noticed, the presence of response bias and central tendency bias can affect the validity and reliability of the use of the Likert scale instrument.

Magdolen et al. (2024) addressed the challenge of ensuring response quality when using item sets with Likert scales in travel surveys. They found a lack of having a universal indicator, they applied several different indicators, compared the results, and developed a new indicator based on correlations which encounter plausible straight lining.

In the field of Kansei Engineering, several exceptional ideas have also been proposed. Shutte and Eklund (2010) introduced an online platform for the Kansei engineering data collection step. They used VAS instead of the Likert scale in the semantic differential method. Betella and Verschure (2016) proposed a modified version of the Self-Assessment Manikin (SAM) rating scale which is called the Affective Slider (AS). AS is a digital self-reporting tool composed of two slider controls for the quick assessment of pleasure and arousal. Fernandes and Yamanaka (2020) presented the Affective Impressions Scale (AIS), a dedicated tool to measure the Affective Accuracy and Affective Synchrony levels of peers in a group.

One of the principle tenets in constructing instruments is that items be as clear and concise as possible. The more items are characterized as cognitively complex, the more likely respondents are to misunderstand the question and answer incorrectly. Even small differences in wording can increase the level of cognitive noise and dramatically alter response patterns (Hodge & Gillespie, 2003).

In practice, most participants in Likert surveys cluster to the extremes. The distribution of responses displays a characteristic *W* shape, with most participants clustering toward the extremes, some expressing indifference, and few in between. Most researchers agree that the W shape does not display the true distribution of preferences, which in reality is most likely a bell-shaped or normal curve (Posner & Weyl, 2018). Jacoby and Matel (1971) mentioned: *Based upon the evidence adduced thus far, reliability should not be a factor in determining a Likert-type scale rating format, because it is independent of the number of scale steps employed*. They compared the results of a survey in three-steps and five-steps Likert scales. There was no significant difference in the results. The result of that experiment supports Posner's claim.

In order to compare the discrete semantic differential with another rating scale, the Visual Analog Rating Scale (VAS) is selected, since VAS is applied to a considerable number of KE research and Dourado et al. (2021) reported the VAS could distinguish non-moderate pleasant emotions induced in observers by images and photos, in comparison to Likert.

The focus of this paper is the second aforementioned problem and the aim is comparing the results of the VAS and Likert. If there was a significant difference in the distributions between the two groups, we propose that researchers use the semantic differential method with a VAS rating scale.

Material and Methods

The case study of the research is reception chairs. We assumed every reception chair can be imagined by three items: backrest, seat, and arm (Figure 1). The items are selected based on the features table for this type of chair, in an online shopping platform. For every item some categories are assumed which are presented in Table 1, these categories are the most repeated categories for each item in the online shopping platform.

Figure 1: *Three items for the current research: Backrest, Seat, Arm.*

As presented in Table 1; two categories for the backrest, three categories for the seat, and two categories for the arm are considered for the current research. It means 12 samples (2*3*2 = 12) must be presented to observers. As it is noted in the first section, samples must be selected from the already existed products in the market. Three combinations (three samples) of the mentioned categories could not be found in Iran's online market (not found samples: b2a1s1, b2a1s3, b2a2 s3). It seems that chairs with a half backrest are not popular in Iran's market and are not offered to the market by manufacturers. Although it impacts the design process in generalizing the results of such KE process, it does not affect the outcome of this research. The presented samples to observers are presented in Table 2.

Table 1: *The list of items and categories.*

#	Item	Categories
1	Backrest	Whole (B1), Half (B2)
2	Arm	With (A1), without (A2)
3	Seat	Plastic (S1), with pillow (S2), with thick pillow (S3)

Through a web engineering technique – it is called scraping - all customers' reviews for this type of chair were fetched from the online shopping platform. Commonly used adjectives were filtered and through the most famous Persian online dictionary (Moin Encyclopedic Dictionary) for every filtered adjective, one contradictory adjective was found. Eventually, eleven paired contradictory adjectives were presented to observers. Table 3 includes the adjectives in Persian and their English translation.

Table 2: *The list of items and categories.*

Marker	Images	Marker	Images	Marker	Images
B1A1S1		B1A2S1		B2A1S2	
B1A1S2		B1A2S2		B2A2S1	
B1A1S3		B1A2S3		B2A2S2	

All the samples and adjectives are presented to the observers through an online platform (www.kansei.ir) which was developed by the researchers for this purpose. The platform has two types of rating scales that are discussed in the paper. Figure 2 is presenting an image of the platform in both situations. Every user filled the form with randomly ordered samples, while the order of adjectives was the same for all users.

Table 3: *Kansei words (adjectives).*

English	Persian
Intimate – Formal	خودمانی - رسمی
Ugly – Beautiful	زشت – زیبا
Punk - Stylish	فکستنی – شیک
Fragile - Resistant	شکننده – مقاوم
Soothing - Stressful	آرامش‌بخش – تنش‌زا
Ergonomic – Not ergonomic	طراحی اصولی – طراحی غیر حرفه‌ای
Modern – Classic	مدرن – کلاسیک
Fake – Original	فیک – اصیل
Clumsy - Elegant	زمخت – ظریف
Decorative - Applicable	تزئینی – کاربردی
Deformed – Eye-catching	بد شکل – چشم نواز

49 participants in two groups filled out the forms. The first group, filled the form with a Likert rating scale (n=26, average of ages = 24, SD=3.3), The second group, filled out another type of form with a VAS rating scale (n=23, average of ages=25, SD =3.8). All participants were design students in Iran universities, which found the forms (the link) in their classroom online forums. The methodology of sampling in the current paper is considered as simple random sampling.

Figure 2: *Left: A sample of the Likert rating scale in Kasnei.ir platform. Right: A sample of the VAS rating scale in Kansei.ir platform.*

Results

Figure 3 presents the distributions of scores in both methods. The VAS scores were categorized into five groups (every group includes 20 consecutive scores). Unlike the semantic differential data, a notable dip was found in the centric category of the VAS graph.

A t-test is done on the average score on products between two categories. The result shows a significant difference between the two groups (p-value < .05).

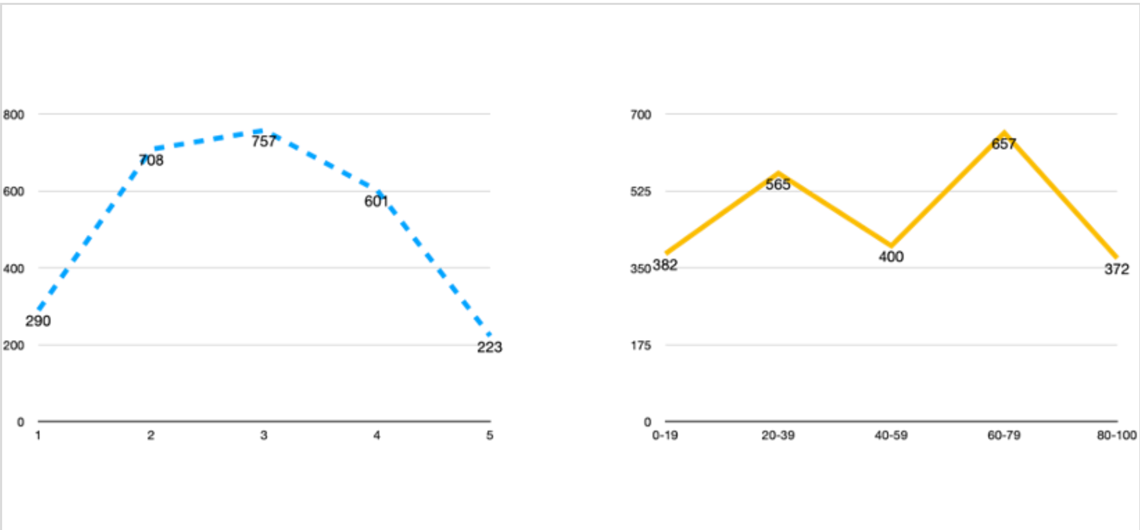
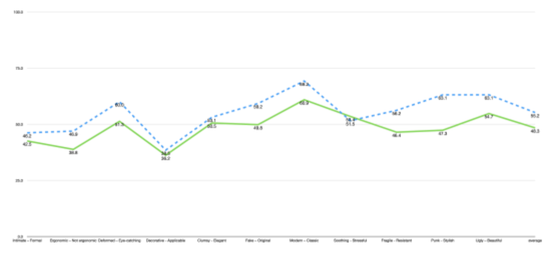
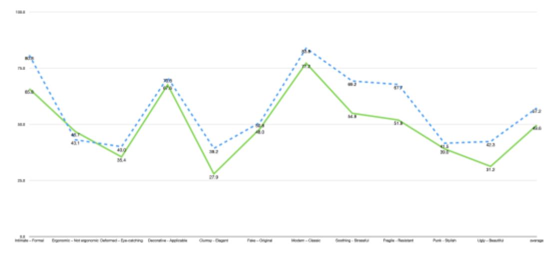
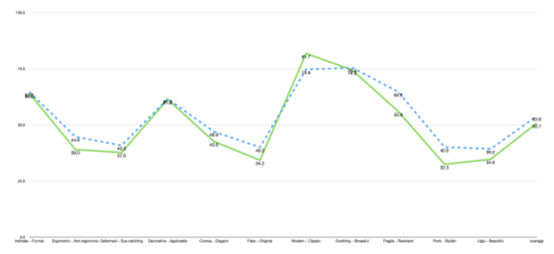
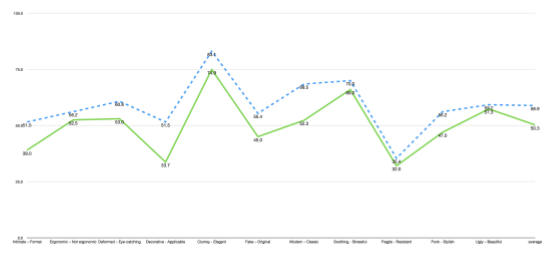
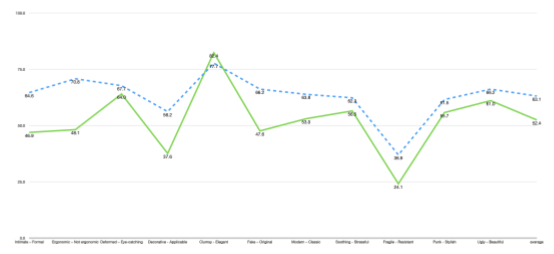
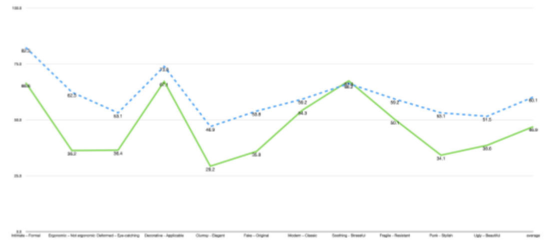


Figure 3: *The left chart represents the distribution of scores using the Likert method. The right chart represents the score using the VAS method. The Y-axis in both graphs represents the frequencies of every score and the X-axis in the right graph represents the scores between 0 and 100 in five clusters.*

Table 4 presents the results of the above-mentioned survey. Each chart represents the average users' score in two different methods, the dashed line shows the result of the Likert method, while the straight line represents the VAS method. The last column of every chart is the average of all scores for the product. The result of the Likert method is scaled from 0-5 to 0-100, In order to display both groups in one graph.

For every item t-test was applied separately and the result of none of them showed a significant difference. This is in contrast to what we observe when running a t-test on the mean of all data at once.

Table 4: *The dashed blue lines represent the result of the Likert method and the green line presents VAS method. The Y-axis shows the average score from 0 to 100. The x-axis is the list of Kansei words. The Kansei words (in order of x-axis): intimate-formal, ergonomic-not ergonomic, deformed-eye-catching, decorative-applicable, Clumsy-elegant, fake-original, modern-classic, soothing-stressful, fragile-resistant, punk-stylish, Ugly-Beautiful, average.*

B1A1S1	B1A2S1
	
B1A2S2	**B1A1S2**
	
B1A1S3	**B1A2S3**
	
B2A2S1	**B2A1S2**
	
B2A2S2	
	

Discussion

As can be seen in Figure 3, the distribution of scores in the Likert method is almost normal, contrary to what Posner and Weyl (2018) that the Likert's survey distribution should be *w-like*. However, it seems that Posner's assumption is based on the gap instinct, the result of the current research shows that Posner's assumption is not true in the field of Kansei Engineering. It is possible that this phenomenon occurs in the field of economics and other behavioral science studies.

The choice of methodology plays a pivotal role in recognizing participants' emotions and the later inference. While Likert scales may introduce a certain level of categorical bias, VAS provides a continuous range that might capture more precise distinctions but could also be inclined to usability biases (these biases occur when observers' responses are shaped by the rating scale rather than their true opinions), as seen in the present study.

In Table 4, the dashed blue lines represent the result of Likert and the straight green lines represent the VAS. In almost all averages of adjectives' scores for each product, the score of the Likert method was greater than in the VAS method, Although the general trends for both rating scales are the same for different adjectives.

The result of the t-test for whole data points of Likert and VAS scores shows a significant difference. We claim that the difference is due to the variance in accuracies between these two methods. It means if we increase the number of Likert inputs to 7 or 9, the accuracy increases, and the difference between the two groups reduces. The significant difference identified by the t-test between the two groups suggests that at least one of the two rating scales is flawed, if not both. Since a definitive baseline cannot be established for such subjective assessments, their accuracy cannot be fully evaluated. However, the discrepancy between the two rating scales indicates that at least one of them contains errors.

Strangely, the distribution chart of the VAS scores is *M-like*. Scores in the range from 40 to 60 were less than expected. We claim the reason for this occurrence is a bias that could have arisen because of the default mode in the survey VAS version. The dip in the central category of the VAS graph may be attributed to the initial presence of a green button in the middle of the scale (In Figure 2, the right image). This visual cue might have influenced users to believe that moving the button was necessary for their input to be recorded. Consequently, even if their opinion aligned with the central category, they might have felt compelled to adjust the button. In future experiments, the green button could be designed to appear only after a user clicks on the scale, thereby mitigating this potential bias.

The study used a relatively small sample size (49 observers) and focused on design students. While this provides valuable insights, it's important to consider how these findings might translate to a broader population. Future studies could explore how scores differ with larger and more diverse participant pools, including non-design students and individuals from different age groups.

The study compared Likert and VAS scales, but other Kansei Engineering studies utilize different methods, such as Fuzzy Measures. Investigating these alternative approaches in future research could reveal additional insights into user perception and potentially identify a method that produces more consistent or nuanced results. For example, Q-sorting, Repertory Grid Technique, Picture Sorting Task.

While the current study utilized images of reception chairs, other studies may require different types of stimuli, such as physical prototypes, virtual reality experiences, or audio recordings. The choice of stimuli should be carefully considered to ensure that it accurately represents the product or concept being evaluated and elicits meaningful responses from participants. Additionally, it is crucial to control for factors such as the order in which stimuli are presented and the context in which they are viewed, as these variables can influence participant responses.

Conclusion

This Kansei Engineering study examined the perception of reception chairs using Likert and VAS rating scales. The findings revealed significant differences between the two rating methods, suggesting that the choice of scale can influence the results. Since the primary objective of this research was not merely to demonstrate differences between the two evaluation methods, as t-test results are evident conducting Kansei engineering using different evaluation methods can yield varying results.

Although it was claimed that the distribution of scores in the Likert surveys is not normal, the result of the current experiment shows the Likert version of the semantic differential is working well in the Kansei Engineering data collection step. In addition, the results of this research suggest the possibility of better accuracy in the semantic differential VAS version in comparison with the Likert version, however, the distribution of the VAS version was not normal.

Future research could explore the impact of sample size, user demographics, different set of Kansei words, and alternative rating scales on study outcomes. Evaluating various rating scales against diverse product categories would provide deeper insights into biases and errors. The M-like distribution of the VAS results can be further explored using an experimental setup that either eliminates default values or incorporates different random values as defaults. Furthermore, investigating the influence of stimuli type and presentation context could provide additional insights into the methodology of Kansei Engineering. By addressing these areas, researchers can enhance the reliability and generalizability of Kansei Engineering studies, contributing to the development of more user-centered and emotionally resonant product designs.

References

Ayas, E. (2011). *Engineering quality feelings*. Doctoral Thesis. KTH Engineering Science. Stockholm, Sweden.

Betella, A., & Verschure, P. F. M. J. (2016). *The affective slider: A digital Self-Assessment scale for the measurement of human emotions*. PLOS ONE, 11(2), e0148037. https://doi.org/10.1371/journal.pone.0148037

Dourado, G. B., Volpato, G. H., De Almeida-Pedrin, R. R., Oltramari, P. V. P., Fernandes, T. M. F., & De Castro Ferreira Conti, A. C. (2021). *Likert scale vs visual analog scale for assessing facial pleasantness*. American Journal of Orthodontics and Dentofacial Orthopedics, 160(6), p. 844–852. https://doi.org/10.1016/j.ajodo.2020.05.024

Fernandes, R., & Yamanaka, T. (2020). *Affective Impressions Scale (AIS): The development of a visual scale to measure affective synchrony among children*. Proceedings of the 8th International Conference on Kansei Engineering and Emotion Research. KEER 2020. Advances in Intelligent Systems and Computing, 1256. Springer, Singapore. https://doi.org/10.1007/978-981-15-7801-4_5

Hodge, D. R., & Gillespie, D. (2003). *Phrase completions: An alternative to Likert scales*. Social Work Research, 27(1), p. 45–55. https://doi.org/10.1093/swr/27.1.45

Magdolen, M., Von Behren, S., Vallée, J., Chlond, B., & Vortisch, P. (2024). *Response bias in Likert-style psychological items – an example from a large-scale travel survey in China*. Transportation Research Procedia, 76, p. 349–360. https://doi.org/10.1016/j.trpro.2023.12.060

Nagamachi, M. (1995). *Kansei Engineering: A new ergonomic consumer-oriented technology for product development*. International Journal of Industrial Ergonomics, 15(1), p. 3–11. https://doi.org/10.1016/0169-8141(94)00052-5

Nagamachi, M. (1999). *Kansei Engineering and its applications in automotive design*. SAE Technical Papers on CD-ROM/SAE Technical Paper Series. https://doi.org/10.4271/1999-01-1265

Nagamachi, M. (2010). *Kansei/Affective Engineering (Systems Innovation Book Series)* [1st ed.]. CRC Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning (No. 47)*. University of Illinois press.

Posner, E. A., & Weyl, G. E. (2018). *Radical markets: Uprooting capitalism and democracy for a just society*. Princeton University Press.

Shutte, S., & Eklund, J. (2010). *Ratingscales in Kansei Engineering: Modification for a European context*. International Conference on Kansei Engineering and Emotion Research 2010.

Hsu, C., Fann, S., & Chuang, M. (2017). *Relationship between eye fixation patterns and Kansei evaluation of 3D chair forms*. Displays, 50, p. 21–34. https://doi.org/10.1016/j.displa.2017.09.002

Jacoby, J., & Matel, M. S. (1971). *Three-point Likert scales are good enough.* Journal of Marketing Research, 8, p. 495–500.

Jebb, A. T., Ng, V., & Tay, L. (2021). *A review of key Likert scale development advances: 1995–2019*. Frontiers in psychology, 12, 637547.

Kreitchmann R. S., Abad F. J., Ponsoda V., Nieto M. D., & Morillo, D. (2019). *Controlling for response biases in self-report scales: Forced-Choice vs. Psychometric Modeling of Likert Items*. Front. Psycho, 10, 2309. https://10.3389/fpsyg.2019.02309

Kusmaryono, I., Wijayanti, D., & Maharani, H. R. (2022). *Number of response options, reliability, validity, and potential bias in the use of the Likert scale education and social science research: A literature review*. International Journal of Educational Methodology, 8(4), p. 625-637. https://doi.org/10.12973/ijem.8.4.625

Tama, I. P., Azlia, W., & Hardiningtyas, D. (2015). *Development of customer-oriented product design using Kansei Engineering and Kano model: Case study of ceramic souvenir*. Procedia Manufacturing, 4, p. 328–335. https://doi.org/10.1016/j.promfg.2015.11.048

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

Zhou, C., Jiang, L., & Kaner, J. (2023). *Study on imagery modeling of electric recliner chair: Based on combined GRA and Kansei Engineering*. Applied Sciences, 13(24), 13345. https://doi.org/10.3390/app132413345