Food Quality and Preference 138 (2026) 105833



Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual





Selecting segmentation strategies for sensory acceptance: Evidence from multiple full-sample and reference-based methods using absolute liking and relative satisfaction

Hae-Jung Park [a,1], Yeon-Joo Lee [a,1], Hee-Eun Lee [a], Yoon-Ah Jang [b], Hye-Seong Lee [a,*]

[a] Department of Food Science and Biotechnology, College of Engineering, Ewha Womans University, Seoul 03760, South Korea
[b] Vegetable Research Division, National Institute of Horticultural and Herbal Science, Rural Development Administration, Wanju-Gun 55365, South Korea

ARTICLE INFO

Keywords:
Absolute liking
Relative satisfaction
DOSD
Reference-based segmentation
Affect magnitude profiling (AMP)
Product testing
Lettuce

ABSTRACT

As the global fresh-cut market grows, meeting diverse preferences is critical to product design and positioning. In practice, sensory acceptance can be measured as absolute liking or as reference-anchored relative satisfaction (e.g., DOSD); the two provide complementary views. This methods-focused study structured segmentation choices by measurement and data transformation, and compared three pathways on a single dataset: (i) segmentation by clustering on absolute liking (9-point; raw and centered 9-point) using AHC (with CLV and K + 1), (ii) reference-based segmentation from DOSD relative satisfaction (*C* summarizing the reference-relative satisfaction difference plus any reference-side bias, d' reflecting bias-free satisfaction discrimination), and (iii) a hybrid that clusters reference-anchored DOSD profiles. Acceptance data ($N = 110$) were collected for eight lettuce cultivars (Korean and European types). In parallel, affect magnitude profiling (AMP) implemented with the double-faced applicability (DFA) test produced d'_A sensory profiles; these were linked to acceptance via PLSR, and segment-specific drivers are presented as interpretable d'_A profiles—consumer-term, SDT-based measures that quantify the magnitude of attribute applicability on a semantic-differential (bipolar) scale. Across pathways, absolute-liking clustering highlighted the overall liking structure; DOSD-based segmentation grouped consumers by satisfaction direction relative to the reference and was less sensitive to scale use; the hybrid combined anchoring with clustering to reveal profile-based subgroups. By explicitly detailing the response-bias implications of each pathway (scale use for liking; criterion/anchoring for DOSD), the study provided a practical decision framework for selecting full-sample (absolute), reference-based (DOSD), or hybrid approaches based on study goals and constraints, thereby supporting segmentation method selection and interpretation.

1. Introduction

In recent years, sensory quality—including how consumers perceive products and which sensory properties they value—has become a key focus for enhancing the competitiveness of new cultivars in fresh-cut product markets (Bowen & Grygorczyk, 2021). Understanding consumer preference and its determinants is critical for market orientation and consumer-driven product development (Grunert et al., 2000; Jaeger et al., 2003). Consumer responses are shaped not only by sensory attributes but also by personal experiences and eating context (Costell et al., 2010), which contributes to consumer heterogeneity. Accordingly, establishing segmentation approaches grounded in reliable acceptance measurement is essential for determining product quality and achieving marketplace success (Bowen et al., 2019; MacFie, 2016; Onwezen et al., 2012; Schröder, 2003).

Consumer segmentation is widely used in marketing to address diverse preferences, understand needs, and guide product development (Goyat, 2011; Lichters et al., 2021; Onwezen et al., 2012). This extends to fresh-cut products, where segmentation helps secure competitive advantage for new cultivars (Baselice et al., 2017; Onwezen & Bartels, 2011). In preference data analysis, cluster analysis has long grouped consumers by response patterns rather than demographics (Ennis et al., 1982). Within food sensory analysis, agglomerative hierarchical clustering (AHC) was commonly applied to hedonic data (Schilling &

* Corresponding author.
E-mail address: hlee@ewha.ac.kr (H.-S. Lee).
[1] ***These authors contributed equally to this work***

https://doi.org/10.1016/j.foodqual.2025.105833
Received 14 January 2025; Received in revised form 3 December 2025; Accepted 9 December 2025
Available online 10 December 2025
0950-3293/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Coggins, 2007), with applications in tomato (Serrano-Megías & López-Nicolás, 2006), orange juice (Varela, 2013), sheep meat (de Andrade et al., 2018), and kale (Swegarden et al., 2019).

Comparative evaluations of clustering and segmentation methods often emphasize internal validity (e.g., within-segment homogeneity and between-segment separation), stability/replicability under resampling or preprocessing choices, and sometimes external predictiveness via links to sensory profiles (Endrizzi et al., 2014; Müller & Hamm, 2014; Wajrock et al., 2008; Westad et al., 2004). In practice, however, researchers choose among methods not only because of such indices but also because methods measure different things and transform the data differently, and these differences can be usefully complementary.

- Hedonic scaling (9-point absolute liking) directly captures overall hedonic level in a familiar, efficient format that scales well to extensive studies and supports full-sample segmentation. Because individuals differ in scale use (e.g., center *vs*. extreme use), analyses often consider both raw and centered 9-point data to focus on preference patterns rather than individual scoring tendencies; this practice is adopted here following the study of Lee et al. (2025) and common guidance (e.g., Vigneau & Qannari, 2002).
- Relative satisfaction collected in reference-paired evaluations (e.g., Degree of satisfaction-difference (DOSD)) is anchored to a reference, providing a complementary lens grounded in signal-detection theory (SDT): C summarizes the reference-relative satisfaction difference plus any reference-side bias, and d' reflects bias-free ("true") satisfaction discrimination between samples (Kim, van Hout, Dessirier, & Lee, 2018; Kim, van Hout, & Lee, 2018). Reference anchoring can help stabilize evaluative criteria by moderating contrast effects while engaging natural perception (Jeong et al., 2016; Kim et al., 2022). As a result, the relative-to-reference matrix can reveal genuine satisfaction-difference patterns that complement what is seen in absolute liking (Kim et al., 2019; Kim, van Hout, Dessirier, & Lee, 2018; Lee et al., 2025).

Methodological developments also broaden the toolbox. Traditional clustering typically assigns all consumers to segments—even when some do not align with a coherent pattern (Vigneau et al., 2016). To address this, Dave (1991) introduced a noise cluster, and Vigneau et al. (2016) embedded this K + 1 strategy within clustering around latent variables (CLV). Unlike AHC, CLV clusters consumers based on correlations among preference vectors (Vigneau & Qannari, 2003); Vigneau et al. (2015) showed that CLV can provide stable, interpretable segmentations relative to k-means, and recommended K + 1 when the aim is to identify segments while mitigating noise. When a benchmark product is clearly defined, reference-based segmentation using attribute-specified DOSD can classify consumers by how they relate to the reference (e.g., satisfaction with thickness, dissatisfaction with too little or too much thickness, or neutrality; Kim et al., 2019)—a framework that complements the full-sample hedonic approach.

The present study, therefore, organized commonly used segmentation families by what they measure and how they transform the data, showing—on the same dataset—what each pathway yields for different research purposes. Three complementary pathways were compared:

(i) segmentation by clustering on absolute liking (9-point ratings, using both raw and centered 9-point to address scale-use differences),

(ii) reference-based segmentation from relative satisfaction using DOSD (C capturing the reference-relative satisfaction difference plus any reference-side bias, and d' capturing bias-free ("true") satisfaction discrimination), and

(iii) a hybrid pathway that applies clustering to reference-anchored DOSD profiles obtained from reference-paired evaluations.

Across pathways, segment solutions were connected to sample-level interpretation using AMP-derived d'_A sensory profiles and partial least squares regression (PLSR) to identify drivers of acceptance, making explicit the response-bias implications of each pathway—i.e., scale-use tendencies for absolute liking (addressed by centering) and criterion/anchoring effects for DOSD (summarized by C and separated from sensitivity via d'). Rather than ranking methods, the objective was to offer a practical decision framework that helps researchers select an appropriate method for their purpose, recognizing that hedonic scaling and DOSD provide complementary information about consumer acceptance.

2. Materials and methods

2.1. Overall experimental design

The overall experimental design, including consumer sensory profiling (Experiment 1) and consumer sensory acceptance evaluation (Experiment 2), is presented in Fig. 1. The same eight lettuce varieties were tested across experiments in two 4-sample sets on separate days (Fig. 2).

In Experiment 1 for consumer sensory profiling, 14 consumer panelists completed the affect magnitude profiling (AMP) test (Fig. 1-①). The procedure comprised four familiarization sessions to develop semantic-differential attributes, followed by the main test, including the two-step rating-based 'double-faced applicability (DFA)' test, administered in three repetitions per sample set (Kim et al., 2026). This produced affect-magnitude d-prime (d'_A) sensory profiles that quantify the applicability magnitude of each attribute for each sample.

In Experiment 2 for sensory acceptance evaluation, 110 consumers completed the Cognitive Reflection Test (CRT) online to assess individual thinking style (Frederick, 2005; Toplak et al., 2014). They then participated in two consecutive acceptance sessions: absolute liking test using a 9-point hedonic scaling with a dummy primer, followed by a relative satisfaction test using the DOSD method with a fixed reference sample (Kim, van Hout, Dessirier, & Lee, 2018) (Fig. 1-② & ③). As summarized in Fig. 3, these two sensory acceptance measurement methods differed in sample presentation format, scaling structure, and output measures, which can produce distinct clustering/segmentation outcomes and different levels of affective discriminability across samples.

This study compared segmentation strategies on a single dataset by grouping commonly used approaches according to what they measure and how they transform the data, and then evaluating three complementary pathways: (i) Clustering on absolute liking (9-point), implemented as agglomerative hierarchical clustering (AHC) on raw ratings and as AHC, clustering around latent variables (CLV), and CLV with K + 1 on centered ratings to reduce scale-use heterogeneity; (ii) reference-based segmentation of relative satisfaction (DOSD), using the individual C to capture reference-relative satisfaction differences (including any reference-side bias) and the d' as a bias-free measure of "true" satisfaction discrimination; and (iii) hybrid (AHC on reference-anchored DOSD ratings obtained from reference-paired evaluations). This design permitted a direct, side-by-side comparison of segmentation families across methods and integrates sensory profiling with acceptance outcomes to reveal cluster-specific sensory drivers. Complete procedural details and data-collection steps are provided in Section 2.4.

2.2. Samples

Eight lettuce varieties were used, representing four salad-use cultivars: leaf, romaine, butterhead, and multi-leaf. Within each cultivar, two widely consumed varieties were selected: Hacheong (HC) and Galmaebit (GM) for leaf; Mansang (MS) and Romaine (RM) for romaine; Sunred-butter (SR) and Redstarsaladball (RS) for butterhead; and Ezabel (EB) and Ezatrix (ET) for multi-leaf. Among the various varieties of Korean leaf lettuce, the HC variety was selected as the reference sample in this study due to its suitability for salad use (Kim et al., 2026).

All the lettuce samples were cultivated at the National Institute of Horticultural and Herbal Science (NIHHS) of the Rural Development

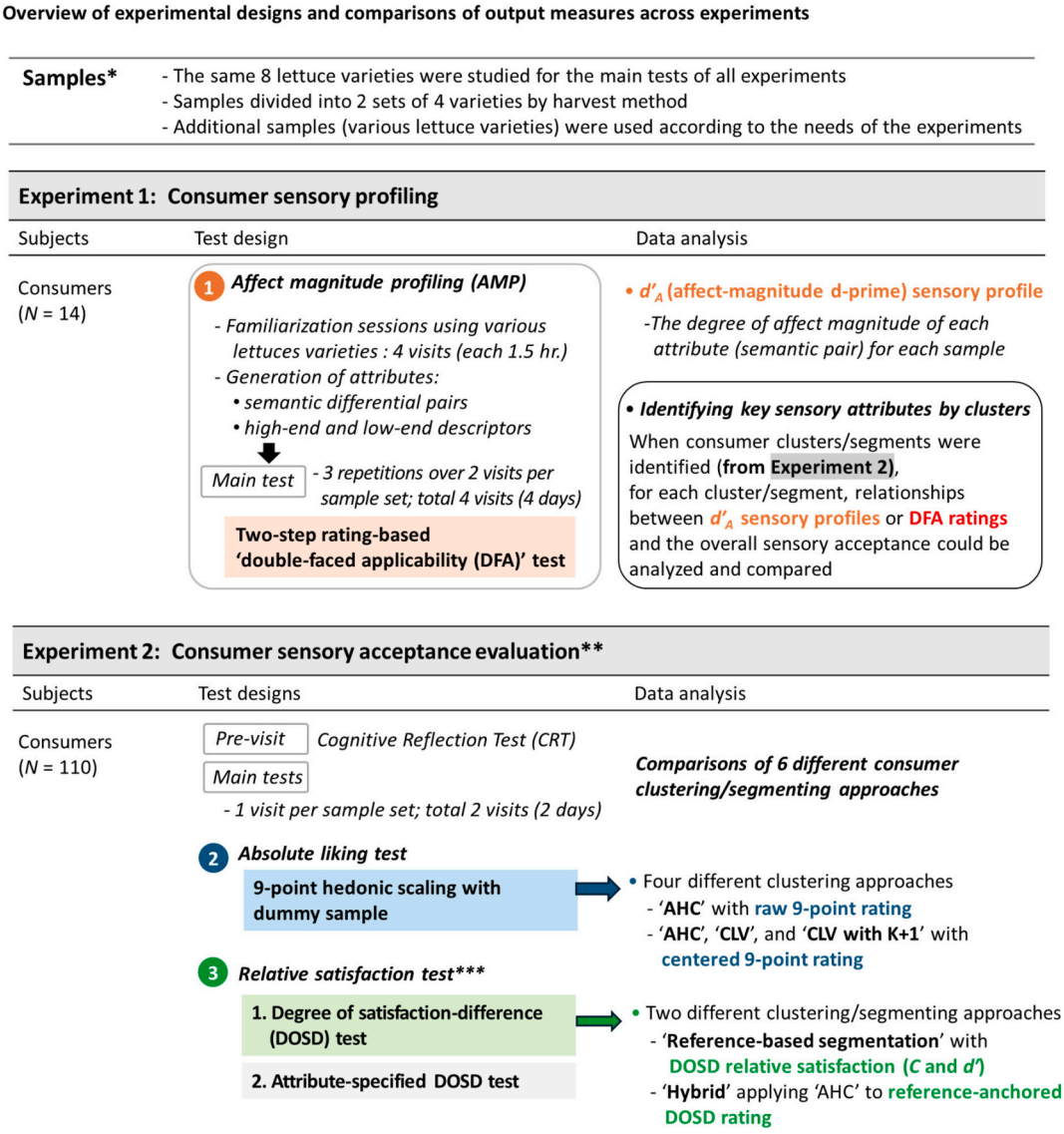


Fig. 1. Overview of two experiments (sensory profiling and sensory acceptance evaluation) and their data analysis pathways.
* Detailed information on the samples is shown in Fig. 2. **Details of the consumer sensory acceptance evaluation—including the test designs and the pathways to consumer clusters/segments—are presented in Fig. 3. ***Data-processing procedures for the relative satisfaction test are provided in Supplementary Fig. 1 (A) and 2 (A).

Administration (Wanju-gun, Jeollabuk-do, South Korea). To control for growth-stage effects on sensory quality (Beaulieu et al., 2004; Shewfelt, 2014), harvest dates and methods were tailored by cultivar. Leaf and romaine lettuces were harvested using a cut-and-come-again approach at eating maturity. In contrast, butterhead and multi-leaf lettuces were harvested once-and-done after semi-head formation. The eight varieties were collected in two harvest methods, each set containing four varieties. Detailed information for each variety is presented in Fig. 2.

Samples were prepared to reflect the context of fresh-cut salad usage. Leaves were cut into bite-sized pieces (8 × 4 cm). To maintain texture consistency across samples, the mid-leaf portion containing both the midrib and the lamina was used, while multi-leaf varieties were served as whole leaves because they were "one-cut-ready". Prepared samples were stored at 4 °C, then equilibrated to room temperature for 30 min before serving. Portions were sealed in zip-lock bags and labeled with three-digit random codes. Portions for the reference sample (HC) were coded with an "R".

To evaluate the intrinsic sensory characteristics, no condiments or accompaniments were provided. Consumer panelists were instructed to consume each portion in one bite. Room-temperature water and unsalted crackers were supplied for palate cleansing between samples.

2.3. Consumers

Participants for the consumer sensory acceptance test and the consumer sensory profiling (AMP) test were recruited via online communities at universities in Seoul, Korea. Eligibility required ages 20–30 years, absence of allergy to lettuce, and self-reported salad consumption of at least three times per month.

For the sensory acceptance test, 110 consumers participated (mean age ± SD = 22.7 ± 2.0 years; 20 males, 90 females). Participants were instructed to refrain for at least 1 h before testing from consuming foods or beverages other than water, brushing their teeth, and using strongly scented cosmetics.

For the sensory profiling test (AMP), 14 consumers participated (mean age ± SD = 23.7 ± 2.9 years; 4 males, 10 females). Panelists were screened with a taste-matching task, aroma identification, a descriptive ability test, and a sensory discriminability test of sweet, bitter, sour,

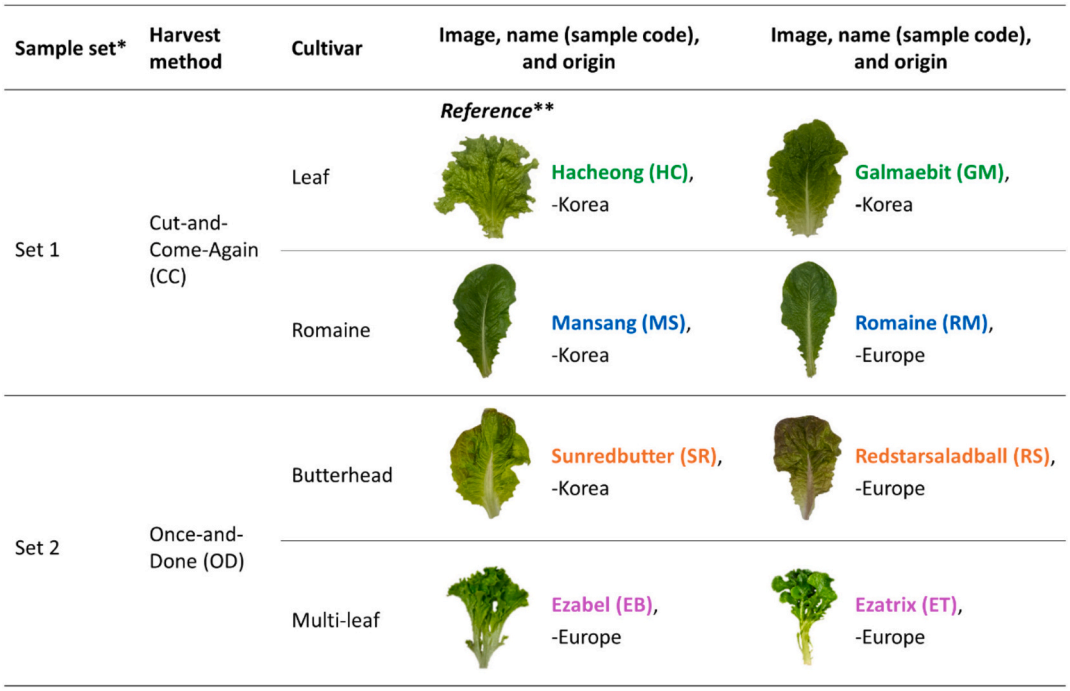


Fig. 2. The eight lettuce varieties used in the experiment.
*Samples were divided into two sets based on harvest method, with each set comprising four lettuce varieties. ** The fixed reference for the DOSD was Hacheong (HC), suitable for salad (Kim et al., 2026), presented with every test sample.

salty, umami, and astringent using a 2-alternative forced choice test (Bakke et al., 2018; Meilgaard et al., 1999). To avoid familiarity with the samples and preserve natural consumer responses in the sensory acceptance test, the sensory profiling panel did not overlap with the acceptance panel.

All procedures were approved by the Institutional Review Board of Ewha Womans University (IRB No. ewha-202312-0021-02). Participation was voluntary, and written informed consent was obtained from all individuals. All participants received monetary compensation for their time.

2.4. Test procedure

2.4.1. Experiment 1: Consumer sensory profiling using affect magnitude profiling (AMP)

AMP was used to derive consumer-based d'_A sensory profiles for the lettuce varieties, following the methods of Chung et al. (2024) and Kim et al. (2026). Unlike previous studies, to stabilize consumers' evaluation criteria, the procedure incorporated a reminder approach (Hautus et al., 2021). In each trial, the reference sample (HC) was evaluated in a fixed-first position and always provided with a test sample to remind consumers of the baseline attribute intensities typical of lettuce.

2.4.1.1. Familiarization session. Four familiarization sessions were conducted twice per week, each lasting 90 min. Consumers were introduced to multiple lettuce varieties to experience and, while tasting, generated sensory attributes with reference to published attribute lists for leafy vegetables (Batziakas et al., 2019; Dinnella et al., 2014; Løkke et al., 2012; O'Brien, Simon, & Tepper, 2022; Talavera-Bianchi et al., 2010). Individually generated attributes were then discussed in group meetings. Consensus was defined as full agreement on the most commonly identified attributes.

For each agreed attribute, paired positive (+) and negative (−) descriptors were formulated in line with consumers' expectations for an ideal lettuce (e.g., 'Not bitter (+) ↔ Bitter (−)'). The final list comprised 16 semantic-differential pairs (Table 1; Supplementary Table 1 (in Korean)). A pre-test familiarized consumers with the attribute list and the two-step response format used in the main test.

2.4.1.2. Main test. The double-faced applicability (DFA) test method was used for AMP (Kim et al., 2026). For each descriptor pair, consumers first answered a binary Yes/No question indicating applicability, then rated the sureness of that response on a 5-point category scale (1 = "not sure", 3 = "sure", 5 = "very sure"). Descriptors were presented in random order within each category. To minimize intentional response shifts, negative and positive descriptors were presented on separate questionnaires (Kim et al., 2017a). The main test comprised three sessions that provided three repetitions per sample. A five-minute rest was given between repetitions to reduce sensory fatigue.

2.4.2. Experiment 2: Consumer sensory acceptance evaluation

2.4.2.1. Cognitive reflection test (CRT) for consumer thinking style evaluation. Before participating in the sensory acceptance test, participants completed an online cognitive reflection test (CRT) to assess their individual thinking style. The 7-item CRT (Toplak et al., 2014), an extension of the original 3-item CRT (Frederick, 2005), was used. Items were designed to appear simple yet require deliberative reasoning to be answered correctly. A 3-min time limit was imposed to standardize testing conditions across participants.

2.4.2.2. Sensory liking test using 9-point hedonic scaling. Consumer sensory liking was measured with a structured 9-point hedonic scale comprising nine labeled categories from "extremely dislike (1)" to "extremely like (9)" (Fig. 3 (A)). After consuming each test sample, participants answered the prompt "How much do you like this sample?". The eight test samples were divided into two sets of four based on harvest method and assessed in two separate sessions (one set per session), with each session conducted on a different day. To minimize tried-first bias, a dummy (primer) sample was presented first in both sessions (Dorado et al., 2016; Kim & Lee, 2015; Lawless & Heymann, 2010; Lee & Meullenet, 2010; Moskowitz et al., 2012). The dummy sample was

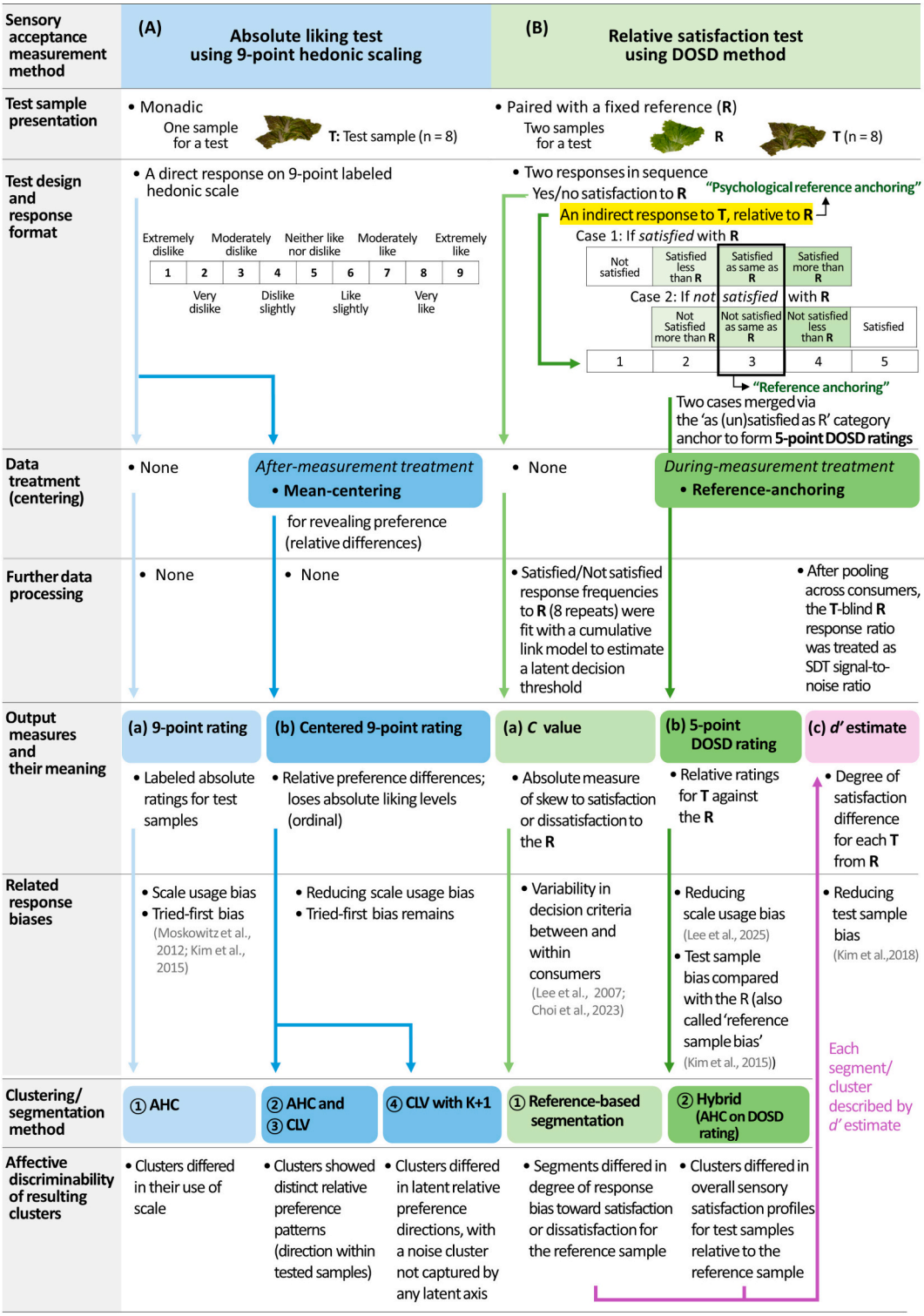


Fig. 3. Comparison of clustering/segmentation pathways: data formats, preprocessing, output measures, and applicable clustering/segmentation approaches.

caipira lettuce (crisphead, Batavia), selected for its common use in salads (Hayes & Simko, 2016). Test samples within each session were presented monadically in a randomized, counterbalanced order to reduce order effects. A 5-min rest followed the hedonic task to limit sensory fatigue before commencing the sensory satisfaction test.

2.4.2.3. Overall sensory satisfaction test using the degree of satisfaction-difference (DOSD) method. The DOSD method followed the procedures of Kim, van Hout, Dessirier, and Lee (2018). The question flow and response options are illustrated in Supplementary Fig. 1 (A). DOSD extends the A–Not A with reminder format by explicitly assessing both the fixed reference and the test sample within each trial. Thus, the reference sample (HC) was always paired with a blind test sample and presented in the first position. The reference thus serves as a psychological anchor at the time of judgment.

In each pair of samples, consumers first evaluated the overall sensory satisfaction of the reference, then judged the test sample relative to it.

Table 1
Affect Magnitude Profiling (AMP)-derived attributes and bipolar descriptor pairs (semantic-differential, consumer terms).

Category of sensory attributes	Pair of descriptors for each attribute: Bipolar descriptors (−/+)		Definition
	Negative descriptors (−)	Positive descriptors (+)	
Appearance	Light green color	Deep green color	The intensity of the deep green color in the leaf, excluding the stem
	Multicolor	Unicolor	Consistency of color in the leaf, excluding the stem
	Not Fresh	Fresh	The intensity of leaf damage and wilting
	Sleek	Crinkled	The intensity of crinkling on the leaf surface
Integrated taste/flavor	Sweet	Not Sweet	The overall sweetness of vegetables
	Bitter	Not bitter	The overall bitterness of vegetables
	Strong green flavor	Weak green flavor	The intensity of the grassy flavor
Texture/ mouthfeel	Not crispy	Crispy	The crispness of the lettuce
	Astringent	Not astringent	The intensity of the astringency in the lettuce
	Stringy	Soft	The ease of chewing the lettuce
	Dry	Moisturized	The degree of moisture released when chewing the lettuce
	Rough	Smooth	The texture felt on the tongue when eating lettuce
Sentimental consequence	Makes me feel unrefreshed	Makes me feel refreshed	The degree of health and refreshing feeling after eating lettuce
	Not taste good	Taste good	The degree of satisfaction with the taste of lettuce
Way of eating (usage)	Not suitable for salad	Suitable for salad	The degree of suitability of lettuce for use as a salad
	Not suitable for being mixed with seasoning	Suitable for being mixed with seasoning	The degree of suitability of lettuce for being mixed with seasoning

The reference sample was evaluated in a two-step sequence: (1) a forced choice of "satisfied" or "not satisfied", followed by (2) a 4-point sureness rating of the previous response ("very sure", "sure", "unsure", or "very unsure").

Consumers then tested one of the test samples and responded on a scale that combined absolute and comparative satisfaction. Two different scales were used depending on the satisfaction judgment for reference. If the reference was "satisfied", the test scale comprised "not satisfied" and satisfied, each with three comparative levels: "satisfied less than the reference", "satisfied as the reference", or "satisfied more than the reference". If the reference was "not satisfied", the test scale comprised "satisfied" and not satisfied, each subdivided into "not satisfied more than the reference", "not satisfied as the reference", or "not satisfied less than the reference".

2.4.2.4. Texture satisfaction test using the attribute-specified DOSD method. Texture is a key driver of lettuce preference, with crunchiness and moisture identified as salient cues of consumer preference (Ansah et al., 2015; Barrett et al., 2010). Building on the overall DOSD procedure, an attribute-specified DOSD was administered to evaluate texture satisfaction, as illustrated in Supplementary Fig. 1 (B) following Kim et al. (2019).

Immediately after the overall satisfaction task with the DOSD method, consumers assessed the reference sample's texture satisfaction using an absolute judgment with sureness rating: "satisfied, sure", "satisfied, unsure", "not satisfied, sure", or "not satisfied, unsure". If "not satisfied" was selected, consumers indicated the reason for dissatisfaction on a four-category scale: "too much crunch", "too much moisture", "insufficient crunch", or "insufficient moisture".

Consumers then evaluated the test sample's texture satisfaction using the same absolute-with-sureness scale: "satisfied, sure", "satisfied, unsure", "not satisfied, sure", or "not satisfied, unsure". A subsequent comparative question asked which was more satisfying relative to the reference, with three response options: "reference sample", "no difference", or "test sample". Texture satisfaction responses from the attribute-specified DOSD were collected but analyzed in a subsequent study and will be discussed on the following paper.

The overall and texture satisfaction tests each comprised eight sample pairs, with a fixed reference and a blind test sample per pair. As in the liking test, pairs were split into two sets of four based on harvest method and evaluated in two sessions on separate days (one set per session). Within each session, pairs were presented in a randomized, counterbalanced order.

The reference was Hacheong (HC) of the leaf cultivar, selected for its suitability for salad (Kim et al., 2026). The reference (HC) was presented repeatedly as the first sample in each pair, and consumers were informed that this first sample was the reference. To minimize response bias, consumers were instructed to judge each reference serving solely on its current sensory experience, even if they recognized it as the same, and to avoid re-tasting the reference while evaluating the test sample to limit carryover. The reference, HC, was also included as a blind test sample ("blind reference"). Consumers were not informed of this duplication and evaluated it as a regular test sample.

2.5. Data analysis

2.5.1. AMP derived signal detection affect magnitude (d'_A) analysis for sensory profiling of each test sample

2.5.1.1. Preparation of data matrices. The two-step responses for each descriptor in the AMP test were converted to a 10-point rating by combining the first-step Yes/No decision with the second-step 5-point sureness rating (Kim et al., 2017b; Kim et al., 2026). For example, responses for "Yes" and "5 (very sure)" were converted to a rating of "10". Similarly, responses "No" and "5 (very sure)" were converted to a rating of "1".

2.5.1.2. Computation of affect magnitude (d'_A). Within the SDT framework, d'_A estimate was computed by treating 10-point ratings for positive descriptors as the "signal + noise" distribution, and ratings for the corresponding negative descriptors as the "noise" distribution (Kim et al., 2017b; Kim et al., 2026). Because the variance of the response criterion is unspecified (Lee, van Hout, & O'mahony, 2007), nonparametric d'_A estimation was used, which avoids assumptions about subjects' response distributions (Bi et al., 2013; Hautus et al., 2021; Lee & van Hout, 2009). For each paired matrix of "signal + noise" and "noise" responses, the area under the receiver operating characteristic (ROC) curve, denoted as Ag, is estimated by maximum likelihood for paired data (Dorfman & Alf, 1969). Ag was the probability that a randomly selected signal observation exceeds a randomly selected noise observation, providing a direct index of discriminability. It was mathematically equivalent to the R-index (Brown, 1974), a non-parametric index widely applied in sensory and consumer research (Bi et al., 2013). Finally, the resulting R-index (Ag) was transformed to d' using Green's area theorem (Green & Swets, 1966), which links ROC area to the expected

proportion-correct for an unbiased observer in a hypothetical paired comparison task. Estimates of the variance and standard error were obtained using the bootstrap method with a total of 10,000 replications (Bi et al., 2013).

The d'_A estimates (range: − 4 to + 4) represent the valence and applicability of each attribute for each sample. The d'_A estimates significantly higher than zero indicate greater applicability of positive (+) descriptors; d'_A estimates significantly lower than zero indicate greater applicability of negative (−) descriptors; d'_A estimates equal to zero indicate that both or neither pairs of descriptors applied (Kim et al., 2017a; Kim et al., 2026; Kim, van Hout, & Lee, 2018a; Lee et al., 2021). Significant differences among multiple d'_A estimates were tested using "dstest", one of the R/S-Plus codes authorized by Bi (2008).

2.5.2. Data treatment for absolute liking

For each of the eight lettuce samples, mean liking ratings were computed within each consumer cluster by averaging 9-point liking responses (Fig. 3 (A)-(a)). To mitigate individual differences in scale use, ratings were mean-centered within the consumer by subtracting each consumer's across-sample mean from all of that consumer's ratings (Fig. 3 (A)-(b)). No variance rescaling (e.g., z-scoring) was applied, preserving within-consumer dispersion.

2.5.3. Signal detection-based satisfaction measure from the DOSD method

2.5.3.1. Computation of the degree of response bias toward satisfaction (C value) for reference sample. Per consumer, eight repeated absolute satisfaction judgments for the reference sample, consisting of a binary satisfied/not satisfied response format, were analyzed with a cumulative link model in the R package ordinal (Christensen, 2019) to estimate the decision-criterion parameter C (C value) (Fig. 3 (B)-(a)). The C is an absolute index of response-criterion bias—the extent to which responses tilt toward satisfaction or toward dissatisfaction; values near zero indicate no bias (Kim, van Hout, & Lee, 2018). The estimated C was subsequently used for reference-based segmentation.

2.5.3.2. Transforming DOSD satisfaction data for the test samples into a 5-point DOSD rating. An indirect response of each test sample relative to a fixed reference sample was organized into data matrices following the framework of Kim, van Hout, Dessirier, and Lee (2018). For every test sample—including the blind reference—two data matrix formats were possible: (1) relative-to-reference satisfaction (5-point DOSD rating; Fig. 3 (B)-(b)) and (2) absolute satisfaction. In this study, only the relative-to-reference format was analyzed.

To construct the 5-point DOSD rating (Supplementary Fig. 2(A)), responses from the two conditional scales (given that the reference sample was judged satisfied or not satisfied) were consolidated using the anchor category "as satisfied/unsatisfied as the reference". On this scale, a rating of "3" denotes parity with the reference; ratings < "3" indicate lower satisfaction than the reference, with a rating of "1" meaning that the test sample was not satisfying whereas the reference sample was satisfying. Ratings > "3" indicate higher satisfaction than the reference, with a rating of "5" meaning that the reference sample was not satisfying whereas the test sample was satisfying. Thus, the 5-point DOSD rating encoded both the direction and magnitude of satisfaction differences for each test sample relative to the reference.

2.5.3.3. Computation of the degree of satisfaction difference (d') for each test sample relative to the reference, HC. Relative differences in satisfaction between each test sample and the reference sample (HC) were quantified with d' estimate under the SDT and Thurstonian approaches (SDT Assistant Program, version 2.0) (Fig. 3 (B)-(c)). For each test sample, responses on the 5-point DOSD scale were modeled as a "signal + noise" distribution, whereas responses to the blind reference sample were modeled as a "noise" distribution. This formulation treats the "true scaling bias" associated with responses to the blind reference sample as noise and adjusts the satisfaction pattern for the comparison samples accordingly. The β-strategy model for the equal-variance A-Not A was then fit by maximum likelihood to estimate d' (Kim, van Hout, Dessirier, & Lee, 2018).

In case the data did not fit the model fitting based on the likelihood ratio and the traditional chi-square diagnostics, non-parametric d' estimates were computed, as outlined in Section 2.5.1.2. If the d' estimates were significantly higher (or lower) than zero, it indicated that consumers were more (or less) satisfied with the test sample relative to the reference sample.

2.5.4. Consumer clustering

2.5.4.1. Cognitive reflaction test (CRT) clustering. Consumers' cognitive reflection was operationalized via total CRT scores and used to form two clusters: High Reflection Thinkers (HRT)—those answering ≥ 4 of 7 items correctly—and Low Reflection Thinkers (LRT)—those answering ≤ 3 of 7 (Lee et al., 2025;Yoon, 2022; Yun, 2023). This proportional rule was consistent with earlier studies based on the original 3-item CRT, which defined HRT as ≥ 2 correct and LRT as ≤ 1 correct (Frederick, 2005; Kim et al., 2017a; Kim, Dessirieret, van Hout, & Lee, 2015; Kim, van Hout, Dessirier, & Lee, 2018b). Prior studies have shown that these cutoffs effectively differentiate cognitive styles and are associated with meaningful differences in sensory acceptance between HRT and LRT groups (Yoon, 2022; Yun, 2023). In total, 75 consumers were classified as HRT and 35 as LRT. These clusters were subsequently entered as the consumer thinking style factor in the multi-way ANOVA model.

2.5.4.2. Agglomerative hierarchical clustering (AHC). Agglomerative hierarchical clustering (AHC) was performed with Euclidean distance and Ward's method (XLSTAT software, version 2022, Addinsoft, New York, USA) on three sensory acceptance datasets: (i) raw 9-point ratings (Fig. 3 (A)-(a)-①), (ii) centered 9-point ratings (Fig. 3 (A)-(b)-②), and (iii) 5-point DOSD ratings (Fig. 3 (B)-(b)-②; hybrid).

Clustering the raw 9-point rating—reflecting absolute liking for eight monadically evaluated samples—yielded two clusters, absolute liking with AHC cluster 1 (Liking/AHC-CL 1; $N = 28$) and absolute liking with AHC cluster 2 (Liking/AHC-CL 2; $N = 82$).

When ratings were mean-centered within consumer to reduce scale-use bias and emphasize sample-related differences, AHC produced three clusters: Centered liking with AHC cluster 1 (Centered liking/AHC-CL1; $N = 33$), Centered liking with AHC cluster 2 (Centered liking/AHC-CL2; $N = 35$), and Centered liking with AHC cluster 3 (Centered liking/AHC-CL3; $N = 42$).

Applying AHC to the 5-point DOSD ratings—indexing satisfaction relative to a fixed reference—also resulted in three clusters, hereafter referred to as the hybrid approach (clustering on reference-anchored DOSD ratings): Hybrid cluster 1 (Hybrid-CL 1; $N = 19$), Hybrid cluster 2 (Hybrid-CL 2; $N = 48$), and Hybrid cluster 3 (Hybrid-CL 3; $N = 43$). These cluster memberships were subsequently included as the cluster factor in the multi-way ANOVA model.

2.5.4.3. Clustering of variables around latent variables (CLV). With the centered 9-point ratings, consumers were clustered using clustering of variables around latent variables (CLV) with local groups (positive-correlation clusters), so that only consumers pointing in the same preference direction are grouped (Vigneau et al., 2016) (Fig. 3 (A)-(b)-③). CLV alternates (i) estimating, for each cluster, a latent component proportional to that cluster's mean liking profile and (ii) reassigning consumers by maximizing covariance/correlation with cluster components. The CLV function in the ClustVarLV package for R statistical software (version 4.5.0; R core team, 2025) was used to implement this procedure, with columns centered and the local option enabled. The number of clusters (K) was selected at the distinct jump in the Delta plot (Vigneau

et al., 2015). In this study, the CLV analysis supported three clusters (K = 3) for the centered 9-point data: Centered liking with CLV cluster 1 (Centered liking/CLV-CL1; $N = 43$), Centered liking with CLV cluster 2 (Centered liking/CLV-CL2; $N = 28$), and Centered liking with CLV cluster 3 (Centered liking/CLV-CL3; $N = 39$). These cluster memberships were subsequently included as the cluster factor in the multi-way ANOVA model.

2.5.4.4. Clustering of variables around latent variables (CLV) with K + 1 strategy. To avoid forcing atypical consumers into substantive segments, a K + 1 (noise cluster) strategy was coupled with CLV (Vigneau et al., 2016) (Fig. 3 (A)-(b)-④). Consumers whose correlations with all cluster latent variables fell below a preset threshold q were temporarily assigned to a noise cluster, yielding cleaner and more interpretable clusters. The procedure was implemented in R (version 4.5.0; R Core Team, 2025) using CLV_kmeans with the kplusone strategy in the ClustVarLV package (Vigneau & Chen, 2014). The choice of q governs noise-cluster size: as q approaches 1, membership criteria tighten, and more consumers are routed to noise. Following Vigneau and Chen (2014), $q = 0.3$ was adopted here, a setting at which the substantive clusters showed stable, high H_k (internal homogeneity) and d_k (discrimination ability), while the noise cluster displayed the expected low cohesion and discrimination. As a result, the final CLV with the K + 1 comprised four clusters, including a noise cluster: Centered liking with CLV with K + 1 cluster 1 (Centered liking/CLV (K + 1)-CL1; $N = 31$), Centered liking with CLV with K + 1 cluster 2 (Centered liking/CLV (K + 1)-CL2; $N = 29$), Centered liking with CLV with K + 1 cluster 3 (Centered liking/CLV (K + 1)-CL3; $N = 34$), and noise cluster ($N = 16$). These cluster memberships were subsequently included as the cluster factor in the multi-way ANOVA model.

2.5.4.5. Reference-based segmentation based on the individual C value from the DOSD method. Consumers were segmented according to their decision-criterion bias for the fixed reference sample (HC), as quantified by the individual C estimated from DOSD responses (Fig. 3 (B)-(a)-①). Segmentation was based on whether the individual C differed significantly from zero (95 % confidence interval). This yielded two segments: Reference-based segmentation 1 (Ref-based seg. 1 / $C > 0$; $N = 39$), comprising consumers significantly biased toward satisfaction with HC, and Reference-based segmentation 2 (Ref-based seg. 2 / $C \leq 0$; $N = 71$), comprising consumers not significantly biased toward either satisfaction or dissatisfaction with HC. These cluster memberships were subsequently included as the cluster factor in the multi-way ANOVA model.

2.5.5. Multi-way analysis of variance (ANOVA)

2.5.5.1. Effects of experimental factors on sensory acceptance rating. To evaluate the effects of experimental factors on sensory acceptance, five-way ANOVAs were conducted separately for the absolute liking (9-point ratings) and the DOSD (5-point DOSD ratings). The model included the fixed factors consumer, sample, sample presentation order, harvest date, and consumer thinking style; analyses focused on main effects only. Post-hoc Tukey's HSD tests were applied for pairwise mean comparisons with a significance threshold of $p < 0.05$. All analyses were performed in XLSTAT (version 2022; Addinsoft, New York, USA).

2.5.5.2. Effects of clustering/segmentation methods on sensory acceptance rating. To assess whether consumer clustering accounts for variance in acceptance, ANOVAs were performed separately for each clustering approach (AHC, CLV, CLV with K + 1, and reference-based segmentation) and for each data structure (raw 9-point, centered 9-point, and 5-point DOSD). For a given clustering approach, a two-way ANOVA with factors cluster and sample was applied to the corresponding rating matrix. Tukey's HSD was used for post-hoc pairwise comparisons ($p < 0.05$). All computations were performed in XLSTAT (version 2022; Addinsoft, New York, USA).

2.5.6. Partial least squares regression (PLSR) for identifying key sensory attributes driving consumer sensory acceptance

Partial least squares regression (PLSR) related AMP d'_A sensory profiles of the eight lettuce samples (X) to consumer liking within each cluster (Y). Separate PLSR models were fitted for each clustering approach and for each cluster. The number of latent components was determined by cross-validation. Model fit was summarized by $R^2X(cum)$ and $R^2Y(cum)$, and predictive ability by $Q^2(cum)$ and RMSE from cross-validation (Supplementary Tables 2 and 3). Components were retained up to the most parsimonious latent dimensionality before a clear rise in RMSE, with $Q^2(cum)$ required to be greater than zero to indicate meaningful predictivity (Ling et al., 2023).

For models based on absolute liking data (Supplementary Table 2), predictive performance was consistently adequate: $Q^2(cum) \geq 0.50$, $R^2Y(cum) \geq 0.89$, $R^2X(cum) \geq 0.57$, and RMSE ≤ 0.26. These values indicated a strong explanation of acceptance variation and stable out-of-sample predictivity across clusters derived from AHC, CLV, and CLV with K + 1.

For models based on relative satisfaction (DOSD) data (Supplementary Table 3), fit indices were similarly acceptable, with $Q^2(cum) \geq 0.48$, $R^2Y(cum) \geq 0.70$, $R^2X(cum) \geq 0.36$, and RMSE ≤ 0.29 across both reference-based segments and hybrid (AHC on reference-anchored DOSD ratings). Taken together, the cross-validated metrics supported the use of these PLSR models for interpreting cluster-specific drivers of consumer acceptance. For each model, variable importance in projection (VIP) was used to identify influential sensory attributes; VIP ≥ 1 was interpreted as a key driver of acceptance.

3. Results

3.1. The d'_A sensory profile of salad lettuce derived from affect magnitude profiling (AMP)

The d'_A sensory profiles derived from AMP are shown in Fig. 4. Of the 16 attribute pairs, 15 attribute pairs—all except 'Astringent (–) ↔ Not astringent (+)'—differed significantly among the eight lettuces ($p < 0.05$).

Across the leaf cultivars HC and GM, d'_A sensory profiles were characterized by 'Light green color (–)', 'Unicolor (+)' appearance, 'Crispy (+)' and 'Moisturized (+)' texture, 'Not sweet (+)' taste, a 'Makes me feel refreshed (+)' sentiment, and 'Suitable for salad (+)'. Within this pair, HC was more strongly 'Moisturized (+)', whereas GM showed a 'Sleek (–)' appearance.

The romaine cultivars MS and RM shared 'Unicolor (+)', 'Fresh (+)', and 'Sleek (–)' appearance, and 'Smooth (+)' texture, but diverged in flavor and mouthfeel. MS tended toward 'Deep green color (+)', 'Dry (–)' texture, 'Not sweet (+)' taste, and 'Strong green flavor (–)', while RM was 'Moisturized (+)', 'Not bitter (+)'.

The butterhead cultivars SR and RS were jointly described by 'Deep green color (+)' and 'Multicolor (–)' appearance, 'Soft (+)', 'Moisturized (+)', 'Smooth (+)' texture, 'Weak green flavor (+)', positive sentiment including 'Makes me feel refreshed (+)' and 'Tastes good (+)', and 'Suitable for salad (+)' and 'Suitable for being mixed with seasoning (+)'. Within this pair, SR was 'Not bitter (+)', whereas RS was 'Not sweet (+)'.

Finally, the multi-leaf cultivars EB and ET were commonly perceived as 'Fresh (+)' in appearance and 'Not sweet (+)' in taste. In addition, EB was relatively more 'Crinkled (+)' and 'Crispy (+)' than ET. These d'_A sensory profiles served as interpretable, consumer-term inputs to the PLSR models in Section 3.4, where segment-specific drivers are presented as d'_A profiles.

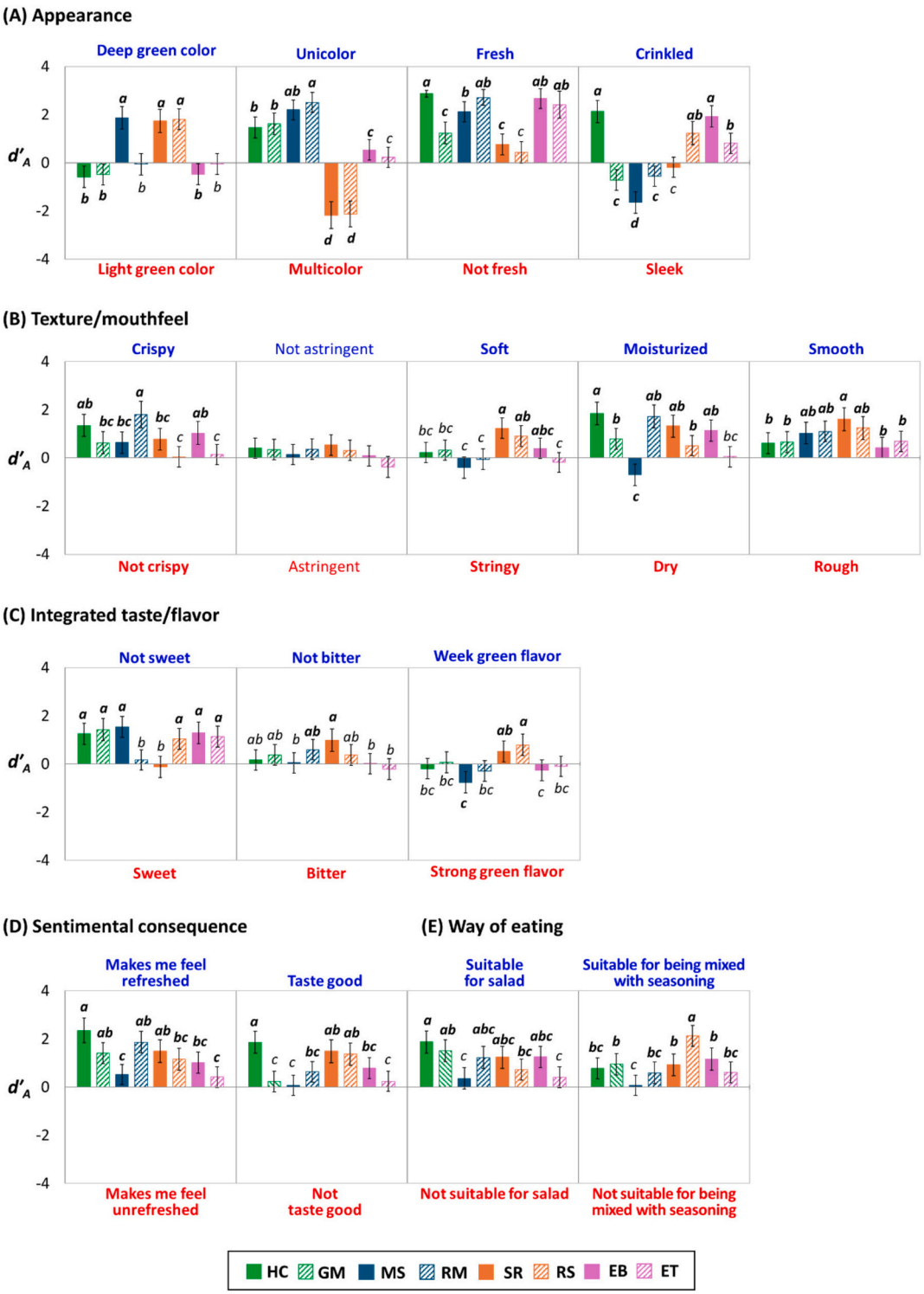


Fig. 4. The d'_A sensory profiles for eight lettuce varieties derived from affect magnitude profiling (AMP). Sensory attributes in bold differ across samples based on the Chi-square test from the dstest of AMP. Within an attribute, different letters denote pairwise differences (95% confidence interval) in d'_A.

3.2. *Effects of experimental factors on sensory acceptance rating*

From the separate five-way ANOVA results performed on the absolute liking (9-point) and DOSD (5-point), *F* statistics are summarized in Table 2. Across both ratings, there were significant main effects of consumer and sample ($p < 0.05$). In contrast, sample presentation order (absolute liking $p = 0.33$; DOSD $p = 0.07$), harvest date (absolute liking $p = 0.92$; DOSD $p = 0.44$), and consumer thinking style (absolute liking $p = 0.45$; DOSD $p = 0.64$) were not significant. Given these results, subsequent analyses focused on between-consumer differences, with thinking style excluded as a primary factor.

3.3. *Examination of clustering approaches on sensory acceptance ratings*

Outcomes across the three complementary pathways were summarized—absolute-liking clustering (raw/centered), reference-based

Table 2
Main-effect ANOVA on absolute liking (9-point) and DOSD (5-point).

Factors	9-point hedonic rating	5-point DOSD rating
Consumer	**3.62**	**1.94**
Sample	**8.65**	**4.44**
Sample presentation order	1.16	2.35
Harvest date	0.09	0.83
Consumer thinking style (HRT/LRT)	0.57	0.22

Values in bold indicate statistical significance ($p < 0.05$).

segmentation from DOSD (*C*), and Hybrid (AHC on reference-anchored DOSD rating)—to align with the framework set in the Introduction. To evaluate how the chosen clustering/segmentation approach influences sensory acceptance outcomes, separate two-way ANOVAs were conducted for each pathway. The resulting *F*-statistics and significance levels are reported in Table 3.

With AHC applied to the raw 9-point rating, the cluster main effect was significant ($p < 0.05$), indicating differences in overall liking levels between clusters, whereas the interaction with sample was not ($p = 0.30$). After mean-centering, the cluster main effect was not significant ($p = 0.27$), but the interaction with sample was significant ($p < 0.05$). The same pattern held for CLV clustering and CLV with K + 1: the cluster

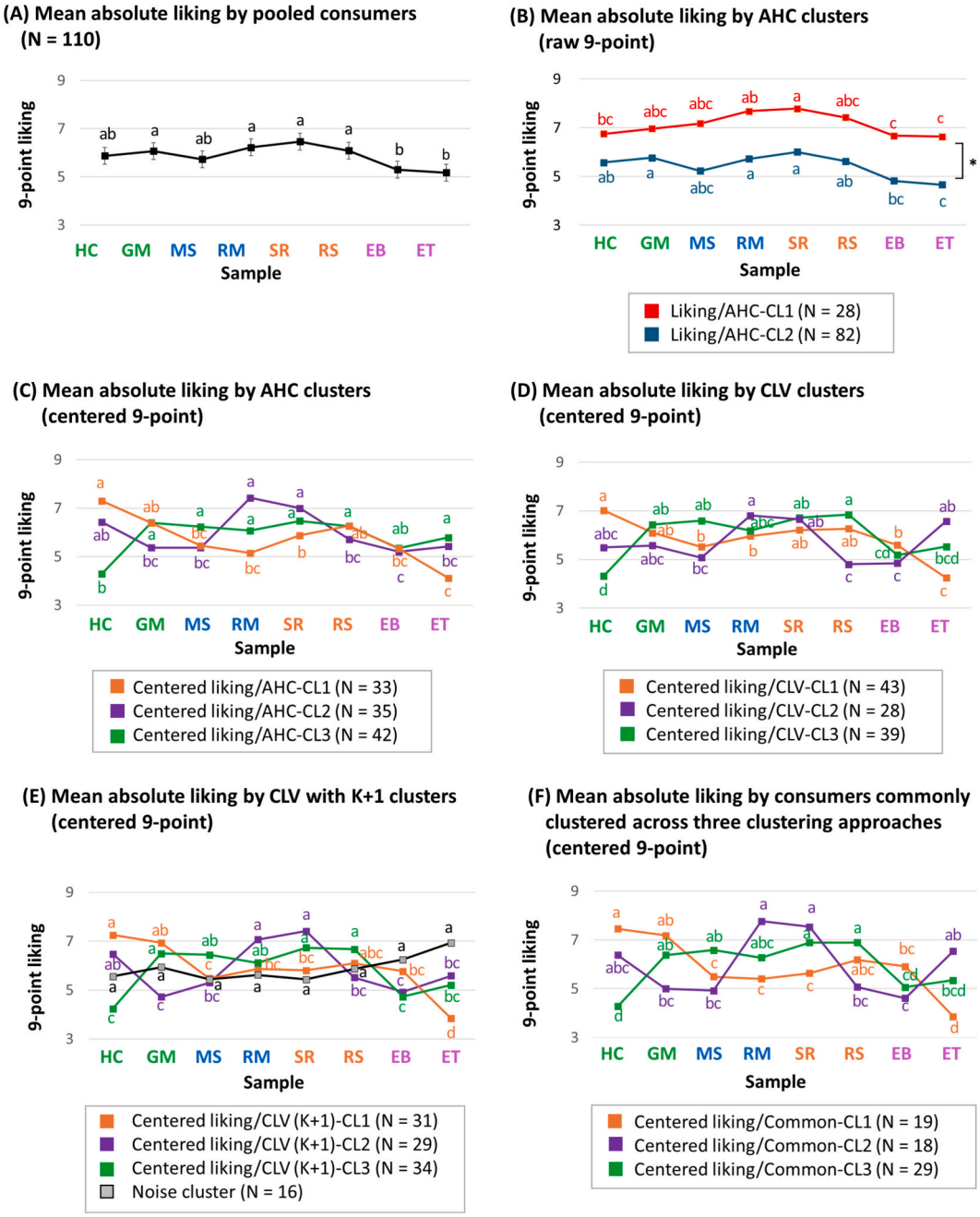


Fig. 5. Cluster-level affective discrimination patterns of samples from absolute liking.
(A) Mean absolute liking by pooled consumers ($N = 110$), (B) Mean absolute liking by AHC clusters (raw 9-point), (C) Mean absolute liking by AHC clusters (centered 9-point), (D) Mean absolute liking by CLV clusters (centered 9-point), (E) Mean absolute liking by CLV with K+1 clusters (centered 9-point), and (F) Mean absolute liking by consumers commonly clustered across three clustering approaches (centered 9-point). Different letters indicate within-cluster differences (Tukey's HSD, $p < 0.05$).

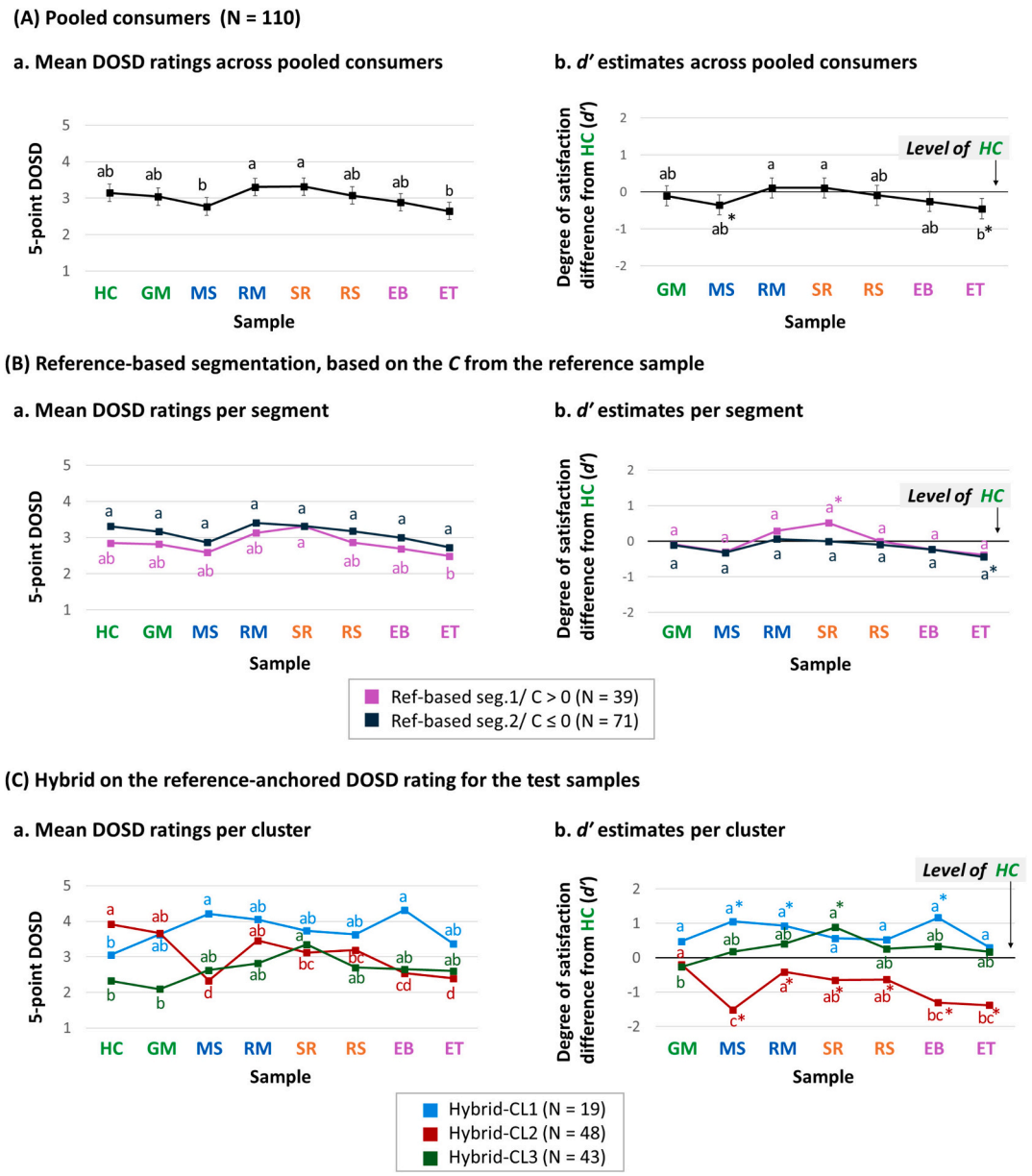


Fig. 6. Segment/Cluster-level affective discrimination patterns of samples from Degree of Satisfaction-Difference (DOSD) relative satisfaction.
(A) Pooled consumers ($N = 110$), (B) Reference-based segmentation by C, and (C) Hybrid (AHC clustering on the reference-anchored DOSD rating) for the test samples. For each (A), (B) and (C), (a) Mean DOSD ratings and (b) d' estimates from the reference sample (HC) for each test sample, by segment/cluster. For (a), different letters indicate within-cluster differences (Tukey's HSD, $p < 0.05$). For (b), different letters indicate 95% confidence interval differences; an asterisk (*) marks estimates significantly different from zero, based on the 95% confidence interval.

main effects were non-significant (CLV $p = 0.26$; CLV with K + 1 $p = 0.91$), while interactions with sample were significant for both ($p < 0.05$). Together, centering reduced scale-usage differences and captured relative preference profiles across samples.

For the 5-point DOSD ratings, applying AHC (hybrid) yielded significant cluster and interaction with sample effects ($p < 0.05$), indicating that hybrid clusters differed in both overall satisfaction and sample-specific preference patterns. In contrast, reference-based segmentation (by individual C) showed a significant cluster main effect ($p < 0.05$) but a non-significant interaction with sample ($p = 0.97$), implying segments that differ in test-sample bias relative to the reference while maintaining a similar ranking of samples across segments. Detailed descriptions and graphical visualizations of these interaction effects are provided in Section 3.4.

3.4. *Affective discriminability and key sensory drivers of acceptance by cluster/segment*

For multiple full-sample clustering on absolute liking, four different cluster-level affective discrimination patterns are shown in Fig. 5. Cluster-specific key sensory drivers were identified by linking AMP d'_A sensory profiles to absolute liking via PLSR; corresponding VIP scores are reported in Supplementary Table 2.

For reference-based segmentation on relative satisfaction, two different segment/cluster-level affective discrimination patterns are presented in Fig. 6. Segment/cluster-specific drivers were derived by linking AMP d'_A profiles to DOSD d' estimates with PLSR; VIP scores are summarized in Supplementary Table 3.

Table 3
Cluster/segment-effect ANOVA on absolute liking (9-point) and DOSD (5-point).

Clustering/segmentation method	Multiple full-sample clustering on absolute liking		Reference-based segmentation on relative satisfaction (DOSD)
Data structure	Raw 9-point	Centered 9-point	5-point DOSD
AHC clustering	**227.74**	1.30	**48.06**
Sample:AHC clustering	1.20	**13.03**	**4.48**
CLV clustering		1.36	N/A
Sample:CLV clustering		**11.23**	N/A
CLV with K + 1 clustering		0.08	N/A
Sample:CLV with K + 1 clustering		**11.47**	N/A
Reference-based segmentation		N/A	**9.76**
Sample:Reference-based segmentation		N/A	0.24

Values in bold indicate statistical significance ($p < 0.05$).
N/A: model not estimated for this data structure.

3.4.1. Absolute liking scores pooled across all consumers

The absolute liking scores for the eight samples, evaluated using a 9-point hedonic scale by a total of 110 consumers, are presented in Fig. 5 (A). Mean liking ranged from 5.17 to 6.46, with significant differences among samples (distinct superscripts, $p < 0.05$). SR (6.46), RM (6.23), RS (6.09), and GM (6.07) constituted the highest-scoring group, whereas EB (5.30) and ET (5.17) were the lowest. HC and MS did not differ significantly from other samples.

3.4.2. Comparison of absolute liking clusters from AHC on raw 9-point

For the absolute liking (raw 9-point), the two AHC clusters differed mainly in scale use: Liking/AHC-CL1 assigned higher scores overall than Liking/AHC-CL2 (cluster means 7.13 *vs*. 5.43). Despite this level shift, their preference patterns were similar across the eight lettuces—both rated SR highest and ET lowest, with the remaining samples following comparable patterns (Fig. 5 (B)). This mirrored the pooled results in Fig. 5 (A), where SR and ET anchored the top and bottom of the distribution, indicating that segmentation revealed heterogeneity in rating level rather than in relative preferences.

PLSR VIP analysis (VIP ≥ 1.0) indicated a shared set of key drivers in both clusters: 'Crinkled (+)' appearance, 'Not bitter (+)', 'Not astringent (+)', and 'Smooth (+)' mouthfeel (Supplementary Table 2). These results indicated that reduced bitterness/astringency and a smooth, crinkled texture were central to acceptance in both clusters. By contrast, attributes related to crispiness and moisture generally showed VIP < 1, suggesting limited impact on liking for either cluster. Overall, hedonic AHC clusters pointed to texture/mouthfeel and the reduction of negative tastes as the principal drivers of increased consumer acceptance.

3.4.3. Comparison of absolute liking clusters from AHC, CLV, and CLV with K + 1 based on centered 9-point

Mean-centering removed scale-use effects and, across AHC (Fig. 5 (C)), CLV (Fig. 5 (D)), and CLV with K + 1 (Fig. 5 (E)), yielded three clusters that expressed relative differences in sensory acceptance across samples. The three approaches produced convergent outcomes in both affective discriminability and key sensory drivers of acceptance. To clearly show cluster-level differences, a homologous subset of 66 consumers consistently assigned to the same cluster across methods was examined (Fig. 5 (F): Centered liking/Common-CL1 $N = 19$, Centered liking/Common-CL2 $N = 18$, Centered liking/Common-CL3 $N = 29$). Relative to the pooled pattern in Fig. 5 (A), clustering revealed pronounced preference reversals rather than a uniform rank order.

For Centered liking/Common-CL1, key drivers were 'Not sweet (+)' taste and mouthfeel ('Not astringent (+)', 'Moisturized (+)'), with added contributions from 'Makes me feel refreshed (+)' and 'Tastes good (+)'. Accordingly, consumers preferred the leaf cultivars HC and GM and least liked ET.

Centered liking/Common-CL2 combined 'Sweet (–)' and 'Not bitter (+)' taste with 'Moisturized (+)' and 'Smooth (+)' mouthfeel, while appearance cues were negligible; this profile aligned with the highest liking for RM and SR and the lowest for EB.

In Centered liking/Common-CL3, drivers shifted toward 'Deep green color (+)' appearance and 'Smooth (+)' mouthfeel, with 'Not bitter (+)' and 'Sweet (–)' playing a more minor role; correspondingly, consumers liked the butterhead cultivars SR and RS and least liked HC.

3.4.4. Relative satisfaction results pooled across all consumers

Relative satisfaction scores from the 5-point DOSD ratings (Fig. 6 (A)-(a)) and the corresponding d' estimates (Fig. 6 (A)-(b)) significantly discriminated among samples ($p < 0.05$). Overall, all samples were rated similarly to the reference (HC), with mean DOSD scores ranging from 2.64 to 3.31. Relatively, RM and SR showed significantly higher satisfaction, whereas MS and ET showed significantly lower satisfaction. Consistently, when calculated as d' relative to HC, MS, and ET were significantly below HC, indicating greater dissatisfaction with these samples.

3.4.5. Comparison of reference-based segmentations derived from C

Consumers were partitioned into two reference-based segments according to satisfaction with the fixed reference HC: Ref-based seg.1 showed a positive criterion toward HC, whereas Ref-based seg.2 did not (Fig. 6 (B)). PLSR VIP results (≥ 1.0) indicated that Ref-based seg.1 was driven by 'Sweet (–)' and 'Not bitter (+)' taste with a 'Smooth (+)', 'Soft (+)', 'Moisturized (+)', 'Not astringent (+)' mouthfeel; appearance was rarely influential (Supplementary Table 3). Consistent with these drivers, SR—sweeter, less bitter, and smoother than HC—was judged more satisfying than HC. Importantly, this segment-level distinction revealed a subgroup that rated SR significantly more satisfying than HC—an effect that was muted in the pooled analysis of all consumers, where no sample was analyzed as more satisfying than HC.

Ref-based seg.2 prioritized 'Sweet (–)' and 'Not bitter (+)' taste together with a 'Crispy (+)' yet 'Moisturized (+)', 'Not astringent (+)' mouthfeel, again with minimal contribution from appearance. In line with these priorities, ET, perceived as less crispy and drier than HC, was less satisfying than HC, while the remaining samples were broadly comparable to HC. Overall, this reference-based segmentation differentiates consumers primarily along a soft/smooth versus crispy texture axis, with both segments favoring reduced bitterness and astringency as central contributors to satisfaction.

3.4.6. Comparison of hybrid clusters obtained by applying AHC to reference-anchored DOSD ratings

Applying AHC to the 5-point DOSD ratings (hybrid) yielded three clusters that differed clearly in their direction of satisfaction relative to the fixed reference HC (Fig. 6 (C)). In Fig. 6 (C)-(a), the cluster profiles represent the mean 5-point DOSD for each test sample and should be interpreted relative to the reference sample (HC). Fig. 6 (C)-(b) shows the corresponding SDT-based results, in which the inherent "true scaling

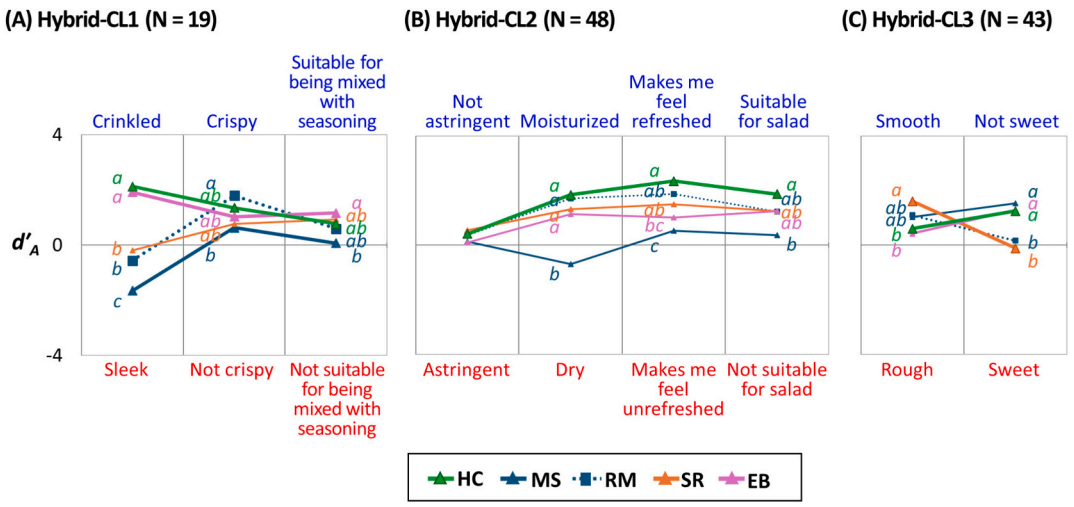


Fig. 7. Examples of AMP d'_A sensory profile for key sensory drivers by hybrid cluster, identified using PLSR on DOSD d' (VIP > 1.2).

bias" associated with the blind reference ratings is treated as noise and used to adjust the satisfaction scores for the other samples, yielding bias-corrected cluster profiles. Summarizing by cluster—as illustrated in Fig. 7 with example contrasts based on PLSR (VIP > 1.2) comparing four lettuces to HC— Hybrid-CL1 (Fig. 7 (A)) placed most significant weight on 'Sleek (–)', 'Not crispy (–)', and 'Not suitable for being mixed with seasoning (–)', while basic tastes ('Sweet (–)', 'Bitter (–)') and 'Astringent (–)' were not influential. Consistent with these drivers, Hybrid-CL1 tended to rate HC relatively low and showed higher satisfaction for MS, RM, and EB.

Hybrid-CL2 (Fig. 7 (B)) emphasized 'Not astringent (+)', supported by 'Moisturized (+)' mouthfeel and positive sentiment ('Makes me feel refreshed (+)') along with 'Suitable for salad (+)'. This profile aligned with higher satisfaction for the reference HC and lower satisfaction for the other samples, most notably MS and ET.

Hybrid-CL3 (Fig. 7 (C)) prioritized a 'Smooth (+)' mouthfeel and 'Sweet (–)' taste, with moisture less influential. In line with these priorities, Hybrid-CL3 showed greater satisfaction for SR than for HC, while the remaining samples were broadly similar to the reference level.

4. Discussion

4.1. Neutralizing sample presentation order bias with a dummy sample in monadic multi-sample hedonic tests

Experimental factors and individual differences can influence responses—especially in multi-sample settings that are vulnerable to presentation-order effects (Lee et al., 2025). Beyond general between-person variability, thinking style, which shapes information processing and decision-making, can modulate how consumers use the 9-point scale in early trials of a monadic sequence. In particular, individuals who rely more on fast, intuitive processing may be more susceptible to contextual anchors and range effects at the start of a session. In contrast, more reflective consumers tend to apply steadier internal criteria (Lee et al., 2025). Presentation-order effects are therefore a persistent threat in monadic hedonic tests because contrast and tried-first (first-serving) biases can shift ratings by position rather than by intrinsic product quality (Lawless & Heymann, 2010; Mazur et al., 2018). Wakeling and MacFie (1995) attributed much of the first-position effect to untrained consumers' initial uncertainty about how to use the hedonic scale, which leads them to treat the first sample differently. These order-related biases can distort sensory acceptance judgments and complicate interpretation.

Counterbalancing mitigates—but does not remove—these influences. Latin-square and Williams designs, along with randomization, ensure that each product is observed equally often in each position and that first-order carryover is balanced across products (MacFie et al., 1989; Wakeling & MacFie, 1995). Such designs improve the reliability of aggregate-level estimates by redistributing position effects. Yet they may not fully protect individual-level analyses. When cluster analysis is applied to segment liking data, clusters are no longer guaranteed to be balanced for sample presentation order. Within a cluster, a particular product could have been observed disproportionately often in the first position. If that product shows higher liking in that cluster, the advantage may reflect presentation order rather than true sensory acceptance. In turn, a cluster that appears distinct could partially be an artifact of presentation order (Hottenstein et al., 2008; Lee et al., 2025).

In Lee et al. (2025), which compared effects of experimental factors on 9-point hedonic ratings versus the reminder-based DOSD method for six cucumber samples, a significant presentation-order effect and its interaction with consumer thinking style were observed for hedonic ratings, despite randomization of presentation order. In particular, both High Reflection Thinkers (HRT) and Low Reflection Thinkers (LRT) gave high ratings to the first-presented sample. LRT, who rely more on fast and intuitive processing, showed greater variability: ratings dropped noticeably for the second sample and then increased gradually from the third sample onward.

Guided by that evidence, the present study implemented a warm-up ("dummy") sample to pre-empt first-position instability. Each consumer first evaluated a mid-range, neutral dummy sample; those data were excluded from analysis (Hottenstein et al., 2008). Mechanistically, the dummy trial (1) absorbs initial scale-use uncertainty among untrained consumers, allowing them to practice mapping sensations to the nine categories before any analyzable trial (Lawless & Heymann, 2010; Wakeling & MacFie, 1995); (2) sets a stable internal criterion so subsequent judgments are made relative to an established yardstick rather than an empty baseline; (3) dissipates novelty-driven contrast/anchoring that is largest at trial 1; and (4) reduces first-trial error variance, which otherwise propagates into cluster-level artifacts when segmenting individual data. Because this dummy-first calibration particularly benefited respondents who are otherwise more order-sensitive (e.g., LRT), the presentation-order main effect on hedonic ratings was not significant in this study, indicating that dedicating the first position to a sacrificial warm-up effectively absorbed the tried-first bias seen previously despite comparable randomization.

The dummy approach is pragmatic but not cost-free. It increases sample usage and session time and can add marginal sensory fatigue in longer flights (Hottenstein et al., 2008). Future work should therefore compare procedural control (dummy-first-sample) with statistical control methods that correct for order post-hoc—e.g., data adjustments

using a first-position offset that subtracts a common value (Hottenstein et al., 2008; Lee & Meullenet, 2010).

4.2. Effects of data treatment (mean-centering and reference-anchoring) on affective discriminability of resulting clusters

Clustering outcomes in consumer testing depend on how effectively data treatment controls response bias. A common remedy is centering, which attenuates between-subject differences in scale use before clustering. Although Baxter (1995) suggested that results from unstandardized and standardized data are often similar in the absence of outliers, and Cherdchu and Chambers IV (2013) reported that standardization may be unnecessary when a single scale type is used, in this study, results showed that—even under a single labeled 9-point scale—centering altered the clustering pattern. This implies that scale-use heterogeneity and effects of experimental factors can still shape partitions, so preprocessing that targets response bias remains consequential.

When hierarchical clustering using Ward's method is applied directly to raw hedonic ratings, the resulting partitions largely reflect between-individual differences in scale use (mean level and range) rather than sample-preference structure. Consequently, clusters separate by overall rating level—how generously or sparingly individuals use the scale (a vertical shift in means)—and remain vulnerable to experimental factors such as presentation-order effects. These level differences are not always mere by-products of response bias; with labeled category scales, they can reflect genuine acceptance-threshold differences and are relevant when the decision target is absolute liking. However, if the aim is to uncover relative preference patterns across samples, relying on raw data can be problematic. As Castura et al. (2023) noted, raw-score–driven partitions may conflate scale use with acceptance and therefore lack portability across datasets, which motivates bias-controlling treatments before clustering.

Once ratings are mean-centered within each consumer, the focus of clustering shifts from who rates high *vs*. low to how consumers differ in their preferences for the samples. Apparent differences in cluster means recede, and informative structure emerges in the cluster-by-sample patterns. In the data of this study, this transition was evident in the ANOVA results: with raw hedonic scores, the cluster main effect was significant while cluster:sample interaction was not; after mean-centering, the cluster main effect disappeared and the cluster:sample interaction became significant—consistently across AHC, CLV, and CLV with K + 1. Thus, mean-centering turned level-driven partitions into pattern-driven ones, improving affective discriminability and enhancing cluster stability against idiosyncratic scale use. These results show that clustering after bias control does more than tidy up scale-use bias—it recovers heterogeneous preference structures that are invisible in pooled means. Pooled means suggested SR high and ET low, whereas centered clusters displayed rank crossings (e.g., a cluster that liked both SR and ET; HC highest in one cluster but lowest in another), revealing actionable subgroups that would have been missed without clustering (cf. Fig. 5 (F)).

This after-measurement treatment (mean-centering) contrasted with the DOSD method's handling of data, which relies on a reference-anchored, during-measurement approach rather than post-hoc correction. In the DOSD task, each response is recorded as a categorical judgment relative to a fixed reference sample, so the evaluative frame is anchored at the moment of satisfaction judgment rather than adjusted afterwards. During data collection, the reference sample thus functions as a psychological anchor: consumers evaluate each blind test sample by explicitly comparing it with the fixed reference, and all responses are made with respect to this benchmark.

After data collection, this reference-anchoring is reinforced by a simple coding rule that normalizes responses around the reference point. Responses indicating "equal satisfaction with the reference" are coded as "3", including cases in which the reference sample is judged satisfying ("Satisfied as same as R") as well as cases in which it is judged unsatisfying ("Not satisfied as same as R"). This procedure effectively centers the data on "parity with the reference", so that subsequent analyses focus on relative differences between samples rather than on absolute satisfaction levels. This procedural reference-anchoring has been shown to reduce scale-use bias (Lee et al., 2025) and, in the present data, produced significant cluster and cluster:sample effects, indicating bias stability without loss of within-cluster, sample-level affective discrimination. Compared with post-hoc mean-centering, the 5-point DOSD rating therefore minimizes additional processing and limits distortion of consumers' original judgments.

Furthermore, in the DOSD-based analyses, Signal Detection Theory (SDT) is applied to calculate d', treating the "true scaling bias" associated with responses to the blind reference sample as noise and adjusting the satisfaction pattern for the comparison samples accordingly. This step functions similarly to traditional centering, in that it removes bias from each blind test sample being evaluated in a pair with the reference sample, thereby allowing fair comparisons between samples.

Taken together, the during-measurement reference-anchoring of the DOSD task, the data preparation, and the SDT-based treatment of blind reference responses collectively perform the role that mean-centering typically plays in hedonic data. The result is a reference-anchored dataset from which clusters can be derived both by consumers' criterion stance toward the reference and by reference-specific satisfaction patterns. In contexts with a credible benchmark, such as the present study, this during-measurement centering provides a coherent path to bias-stable clusters while preserving interpretable differences among products (Lee et al., 2025).

4.3. Comparison of clustering/segmentation methods based on multiple full-sample and reference-based approaches using absolute liking and relative satisfaction

Clustering is widely used to classify consumers with similar response patterns. Consistent with Yenket et al. (2011), it is prudent to compare multiple clustering and segmentation approaches and select the method that matches the dataset and study goals. A comparison of the approaches considered here is summarized in Fig. 8. In sensory and consumer research, multiple full-sample clustering of absolute liking is commonly used in early-stage development, exploratory testing, or when no clear benchmark exists—that is, when the goal is to capture broad preference structures across diverse or unfamiliar product categories (Lee et al., 2025). Two families are most often applied: (i) agglomerative hierarchical clustering (AHC, Ward's linkage), a distance-based, bottom-up procedure that merges clusters to minimize the increase in within-cluster variance and yields a dendrogram; and (ii) Clustering around Latent Variables (CLV), a partitioning approach that learns cluster-specific latent axes (LVs) and iteratively reassigns individuals. The CLV with K + 1 variant adds a noise/overflow group to avoid forcing all consumers into substantive clusters.

Applied to centered 9-point rating, AHC, CLV, and CLV with K + 1 produced similar partitions. After centering within individuals, Euclidean distances between preference profiles behaved much like correlation distance, so both variance-based AHC and latent-axis CLV effectively group by profile shape rather than overall liking level. Moreover, Ward's minimum-variance criterion seeked compact groups in Euclidean space, whereas CLV models cluster-specific latent axes; when a low-rank set of preference directions dominates variation, the methods align with the same leading axes and recover near-identical partitions (Ding & He, 2004; Murtagh & Legendre, 2014; Vigneau et al., 2015).

As the need grows to identify consumers with different preference structures, reference-anchored segmentation offers a more transparent lens than full-sample clustering for absolute liking. Although practice often gravitates toward the "best" and "worst" products for prioritization or diagnosis, many go/no-go decisions depend on parity or superiority relative to a gold standard. Fixing a benchmark focuses inference

Clustering/ segmentation method	(A) Multiple full-sample clustering on absolute liking			(B) Reference-based segmentation on relative satisfaction (DOSD)	
	Raw 9-point	*Centered 9-point*		*C value*	*DOSD rating*
	① AHC	② AHC and ③ CLV	④ CLV with K+1	① Reference-based segmentation	② Hybrid
Practical strengths	• Preserve absolute liking levels for intuitive interpretation • Simple with no data pre-processing • Aligns directly with business pass/fail thresholds	• Reduces scale-use bias; highlights relative preference • (CLV) Cluster consumers by latent preference direction; clearer profiles than distance-only AHC	• Increases cluster purity by parking weakly aligned consumers in a noise cluster • Clearer latent-direction clusters; robust to borderline cases	• Clearer separation by key sensory attributes driving satisfaction • Directly answers 'Who is satisfied with reference?'	• Clusters reveal opposing directions of satisfaction vs. the reference • Reduced sensitivity to scale-use bias; maintains comparability across multi-sample contexts
Limitation	• Cannot tell if cluster differences reflect absolute liking tied to label meaning or from scale-use bias • Susceptible to context/order bias; hard to generalize	• Noisy respondents can dilute cluster purity • Ignores variance • (AHC) extreme raters still influence fit	• Dependent on noise-rate tuning • Very small remaining clusters (after assigning noise) can weaken stability	• Requires repeated reference judgments and a meaningful, stable reference • No absolute satisfaction levels • Sensitive to reference sample stability	• Dependent on the chosen reference • No absolute satisfaction levels
Use when	• Clustering consumers by overall liking patterns across samples • No fixed benchmark/reference is needed • Sample size is moderate-to-large and samples span wide hedonic range			• Segmenting consumers by relative satisfaction vs. a benchmark/reference • Reformulation/quality-upgrade projects require an anchor sample • Needs decisions/claims of 'better/worse vs. reference'	
	• To quantify absolute differences in overall liking • The panel shows uniform/consistent scale use	• To map relative preferences and between-sample discrimination • Strong scale-use heterogeneity among consumers is expected	• To reveal a clear, data-driven cluster structure with distinct preference directions—without forcing all consumers into clusters • Leaving some consumers ambiguous is warranted	• To identify who and how many are satisfied and not satisfied with the reference, rather than how samples are ranked by each segment • Decisions are benchmark-driven (e.g., meet current reference, qualify for claim), and pass/fail vs reference is key	• To form clusters showing directional contrasts (more vs. less satisfying than the reference) across samples • Use when you must screen large sets against a gold standard and need clean, directional contrasts for go/no-go

Fig. 8. Summary of comparisons among clustering/segmentation methods across multiple full-sample and reference-based approaches using absolute liking and relative satisfaction.

on relative satisfaction, revealing how consumers judge products against that reference and complementing level-oriented views based on absolute liking alone.

Evidence converges on this point. By fixing a benchmark and expressing judgments relative to that reference—as in DOSD and affective-reference designs—segments are defined by how consumers evaluate samples against a common standard (Kim et al., 2019; Kim, Sim, & Lee, 2015; Lee et al., 2025). Kim, Sim, and Lee (2015) identified affective discriminators who showed a consistent preference for reference samples—segments with high sensory discriminability—and proposed a consumer-relevant action standard for quality decisions, thereby linking segmentation to practice. Reference-anchored clustering also exposes relative-satisfaction structures that differ from those derived from absolute likings (Lee et al., 2025), and attribute-specified DOSD can target specific sensory attributes of the reference to guide focused reformulation and positioning (Kim et al., 2019). In short, organizing segmentation around a fixed reference integrates statistical robustness with managerial interpretability, providing attribute-level levers that full-sample clustering on absolute liking does not reliably deliver.

Within this framework, two reference-based approaches were applied for lettuce using a DOSD dataset. First, consumers were segmented by their stance toward the repeatedly evaluated reference—either an absolute skew toward satisfaction or toward dissatisfaction. In a chicken-bouillon study with systematic manipulations of mouthfeel attributes (notably thickness and fattiness), Kim et al. (2019) used attribute-specified DOSD to segment consumers by their preferred reference thickness; segments then displayed distinct satisfaction patterns across products spanning different viscosities, consistent with segment-wise differences in perceptual sensitivity. In the lettuce dataset, differences spanned multiple sensory dimensions, so contrasts were less pronounced than under single-axis manipulation; nevertheless, a coherent texture split emerged: consumers satisfied with the reference prioritized soft/smooth texture, whereas consumers neutral to the reference weighted crispiness more strongly. These results support reference-anchored segmentation as a means to reveal benchmark-relative heterogeneity that may be obscured by absolute liking alone, while allowing interpretation to be aligned with category-relevant attributes.

Second, clustering based on relative satisfaction (judgments coded as more/less/as satisfying than the reference) produced segment patterns distinct from those obtained with absolute liking. Associations with sensory profiling showed that relative-satisfaction clusters separated on non-overlapping driver attributes, whereas clusters derived from absolute ratings did not show comparably sharp attribute divisions. This pattern suggests that absolute liking, which lacks a shared benchmark, can blur how consumers compare products, thereby obscuring actionable intervention points.

Overall, the evidence supports the value of reference-anchored

segmentation on relative satisfaction. Anchoring the task to a common reference stabilizes the evaluative frame at the moment of response, limiting scale-use bias and minimizing post-hoc data processing (Kim, van Hout, & Lee, 2018; Lee et al., 2025). The DOSD framework also yields d' estimates grounded in signal detection theory (SDT). In this formulation, responses to the reference are modeled as "noise" and responses to test samples as "signal + noise", producing a probabilistic, group-level account of satisfaction. Importantly, individual perceptual variability—salient for categories with high product-to-product variation, such as fresh produce—is not discarded as random error but absorbed into the model and averaged or counterbalanced in the resulting distributions (Lee et al., 2025). Test sample bias compared with the reference (also called reference sample bias) can also be attenuated (Kim, van Hout, & Lee, 2018c). Together, these features enable robust quantification of whether a test sample is perceived as more or less satisfying than the reference and yield segments that jointly capture criterion stance and sample-specific satisfaction patterns, enhancing explanatory power and decision relevance. In quality-management contexts, reference anchoring further enables actionable thresholds to support go/no-go and equivalence judgments.

Despite these advantages, DOSD is intrinsically reference dependent: its effectiveness hinges on the availability of a meaningful and recognizable benchmark. As noted in Lee et al. (2025), in emerging categories (e.g., cultivated meat, edible insects, novel protein matrices), consumers may lack a stable criterion of 'what satisfying looks like' reducing the interpretability of reference-anchored judgments. DOSD can still be informative when the reference represents a widely accepted standard (house flagship, market leader, or a prototypical sensory target) and when the business question concerns parity or superiority relative to that standard across adjacent line extensions or processing variants. By contrast, categories characterized by high intra-category heterogeneity (no clear prototype) or rapid innovation cycles (shifting targets) are less suitable for DOSD as a primary method; in such cases, response-level centering of absolute liking may provide a more neutral frame.

4.4. Limitations of the present study and future suggestions

All the consumers in this study were in their 20s, as they were recruited through the university community. However, age is known to influence not only overall preferences but also texture preferences for vegetables (Dinnella et al., 2016; Gotow et al., 2023; Poelman & Delahunty, 2011). For instance, Hildebrandt et al. (1997) reported that elderly consumers ($\geq$ 60 years of age) tend to avoid stringy and crunchy foods, such as meat and vegetables, leading to lower consumer preference for texture attributes. This age-related variation underscores the importance of including a more demographically diverse sample in future studies to enhance the reliability and generalizability of the findings.

In the study of Lee et al. (2025), the selection of the reference sample was highlighted as a critical factor affecting the reliability and stability of sensory evaluations, with unsuitable references compromising interpretability. This issue is particularly significant in emerging markets for novel products lacking clear benchmarks. To address this, this study selected the reference sample by utilizing the results of AMP from Kim et al. (2026). By aligning the reference sample with consumer evaluations, this approach ensured it effectively reflected real usage contexts and preferences. However, challenges in reference selection remain, particularly when dealing with highly diverse consumer groups or markets with rapidly evolving preferences. Future research should focus on refining reference selection methodologies, including the integration of dynamic, attribute-specific references and regionally representative sampling. Such efforts will further expand the applicability of the reference-base segmentation approach across diverse product categories and market contexts.

Another promising direction for subsequent research is exploring the relationship between key sensory attribute drivers and metabolite profiles. Currently, the metabolomics approach has gained interest in sensory science, as it provides a comprehensive understanding of the chemical essence of the characteristics involved in flavors, tastes, and textures (Shu et al., 2023; Xie et al., 2021). Furthermore, previous studies have already reported that consumer preference is affected by the metabolite contents of fresh-cut products (Kim et al., 2023; Marsh et al., 2003; Ulrich & Olbricht, 2016). Therefore, identifying the metabolites influencing consumer preferences could serve as a guide for breeding new consumer-preferred cultivars.

5. Conclusion

Three segmentation pathways were compared on the same dataset—clustering on absolute liking (raw and consumer-mean centered), reference-based DOSD, and a DOSD-anchored hybrid—and segment profiles were linked to AMP-DFA d'_A and PLSR-derived acceptance drivers. Segment-specific drivers were reported as interpretable d'_A profiles (consumer-term, SDT-based measures on a bipolar semantic-differential scale), enabling method-agnostic, transparent product interpretation and cross-segment comparability. Absolute-liking clustering summarized absolute preference level across products; DOSD grouped consumers by direction of satisfaction relative to a fixed reference with reduced sensitivity to scale-use; and the hybrid combined anchoring and clustering to reveal interpretable profile-based subgroups. By making the response-bias implications explicit (scale-use for liking; criterion/anchoring for DOSD), a fit-for-purpose framework is provided for selecting full-sample, reference-based, or hybrid approaches according to study aims.

Because DOSD is inherently reference-dependent, its highest utility is expected in categories with a clear benchmark; in categories without a credible reference, absolute-liking clustering or the hybrid approach may be preferable options. In practice, the DOSD task is centered at response—the midpoint ("3") on the 5-point DOSD scale denotes parity with the reference—which attenuates scale-use sensitivity while making direction-of-satisfaction explicit. Together with the SDT-based treatment of blind-reference responses, this reference-anchoring performs the role typically served by mean-centering, ensuring that segmentations reflect true relative satisfaction differences without the need for additional traditional centering of the DOSD ratings. The generalizability of these findings is limited by the relatively young consumer panel. Broader demographic coverage and more systematic procedures for reference selection are warranted.

CRediT authorship contribution statement

Hae-Jung Park: Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Yeon-Joo Lee:** Writing – review & editing, Methodology, Conceptualization. **Hee-Eun Lee:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Yoon-Ah Jang:** Writing – review & editing, Methodology, Conceptualization. **Hye-Seong Lee:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization.

Ethical statement

Ethical approval for the dataset in this study was granted by Ewha Womans University Institutional Review Board Committee, Reference number ewha-202312-0021-02, dtd 10/30/2024.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the 'Cooperative Research Program for Agriculture Science and Technology Development (No. RS-2022-RD010402)', Rural Development Administration, Republic of Korea. The authors would like to give special thanks to Si-Yoon Yang, Min-Kyoung Ko, and Hyun-Jung Cho for their assistance in conducting this experiment.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodqual.2025.105833.

Data availability

Data will be made available on request.

References

de Andrade, J. C., Nalerio, E. S., Giongo, C., de Barcellos, M. D., Ares, G., & Deliza, R. (2018). Consumer sensory and hedonic perception of sheep meat coppa under blind and informed conditions. *Meat Science, 137*, 201–210. https://doi.org/10.1016/j.meatsci.2017.11.026

Ansah, F. A., Amodio, M. L., & Colelli, G. (2015). Evaluation of the impact of nitrous oxide use on quality and shelf life of packaged fresh-cut 'Iceberg'Lettuce and wild rocket. *Chemical Engineering Transactions, 44*, 319–324. https://doi.org/10.3303/CET1544054

Bakke, A. J., Stubbs, C. A., McDowell, E. H., Moding, K. J., Johnson, S. L., & Hayes, J. E. (2018). Mary Poppins was right: Adding small amounts of sugar or salt reduces the bitterness of vegetables. *Appetite, 126*, 90–101. https://doi.org/10.1016/j.appet.2018.03.015

Barrett, D. M., Beaulieu, J. C., & Shewfelt, R. (2010). Color, flavor, texture, and nutritional quality of fresh-cut fruits and vegetables: Desirable levels, instrumental and sensory measurement, and the effects of processing. *Critical Reviews in Food Science and Nutrition, 50*(5), 369–389. https://doi.org/10.1080/10408391003626322

Baselice, A., Colantuoni, F., Lass, D. A., Nardone, G., & Stasi, A. (2017). Trends in EU consumers' attitude towards fresh-cut fruit and vegetables. *Food Quality and Preference, 59*, 87–96. https://doi.org/10.1016/j.foodqual.2017.01.008

Batziakas, K. G., Talavera, M., Swaney-Stueve, M., Rivard, C. L., & Pliakoni, E. D. (2019). Descriptive analysis and consumer acceptability of locally and commercially grown spinach. *Journal of Food Science, 84*(8), 2261–2268. https://doi.org/10.1111/1750-3841.14710

Baxter, M. J. (1995). Standardization and transformation in principal component analysis, with applications to archaeometry. *Journal of the Royal Statistical Society Series C: Applied Statistics, 44*(4), 513–527. https://doi.org/10.2307/2986142

Beaulieu, J. C., Ingram, D. A., Lea, J. M., & Bett-Garber, K. L. (2004). Effect of harvest maturity on the sensory characteristics of fresh-cut cantaloupe. *Journal of Food Science, 69*(7), 250–258. https://doi.org/10.1111/j.1365-2621.2004.tb13624.x

Bi, J. (2008). *Sensory discrimination tests and measurements: Statistical principles, procedures and tables*. John Wiley & Sons.

Bi, J., Lee, H.-S., & O'Mahony, M. (2013). Statistical Analysis of Receiver Operating Characteristic (ROC) Curves for the Ratings of the A-Not A and the Same–Different Methods. *Journal of Sensory Studies, 28*(1), 34–46. https://doi.org/10.1111/joss.12021

Bowen, A. J., Blake, A., Tureček, J., & Amyotte, B. (2019). External preference mapping: A guide for a consumer-driven approach to apple breeding. *Journal of Sensory Studies, 34*(1), Article e12472. https://doi.org/10.1111/joss.12472

Bowen, A. J., & Grygorczyk, A. (2021). Challenges and opportunities for sensory and consumer science in new cultivar development and fresh produce marketing. *Current Opinion in Food Science, 41*, 152–158. https://doi.org/10.1016/j.cofs.2021.04.009

Brown, J. (1974). Recognition assessed by rating and ranking. *British Journal of Psychology, 65*(1), 13–22. https://doi.org/10.1111/j.2044-8295.1974.tb02766.x

Castura, J. C., Meyners, M., Pohjanheimo, T., Varela, P., & Næs, T. (2023). An approach for clustering consumers by their top-box and top-choice responses. *Journal of Sensory Studies, 38*(5), Article e12860. https://doi.org/10.1111/joss.12860

Cherdchu, P., & Chambers, E., IV (2013). Personality classification of consumers: A comparison of variables, standardization and clustering methods. *Journal of Sensory Studies, 28*(6), 504–512. https://doi.org/10.1111/joss.12075

Christensen, R. H. B. (2019). Ordinal—Regression models for ordinal data. R package version 2019.12-10, URL. http://www.cran.r-project.org/package=ordinal/.

Chung, Y.-D., Lim, H.-J., Lee, Y.-J., & Lee, H.-S. (2024). Characterization and quantification of sensory quality attributes of Korean cucumber varieties based on consumer perception and liking. *Korean Journal of Food Science and Technology, 56*(6), 763–774. https://doi.org/10.9721/KJFST.2024.56.6.763

Costell, E., Tárrega, A., & Bayarri, S. (2010). Food acceptance: The role of consumer perception and attitudes. *Chemosensory Perception, 3*, 42–50. https://doi.org/10.1007/s12078-009-9057-1

Dave, R. N. (1991). Characterization and detection of noise in clustering. *Pattern Recognition Letters, 12*(11), 657–664. https://doi.org/10.1016/0167-8655(91)90002-4

Ding, C., & He, X. (2004). *K-means clustering via principal component analysis. In proceedings of the twenty-first international conference on machine learning (p. 29)*. https://doi.org/10.1145/1015330.1015408

Dinnella, C., Morizet, D., Masi, C., Cliceri, D., Depezay, L., Appleton, K. M., & Monteleone, E. (2016). Sensory determinants of stated liking for vegetable names and actual liking for canned vegetables: A cross-country study among European adolescents. *Appetite, 107*, 339–347. https://doi.org/10.1016/j.appet.2016.08.110

Dinnella, C., Torri, L., Caporale, G., & Monteleone, E. (2014). An exploratory study of sensory attributes and consumer traits underlying liking for and perceptions of freshness for ready to eat mixed salad leaves in Italy. *Food Research International, 59*, 108–116. https://doi.org/10.1016/j.foodres.2014.02.009

Dorado, R., Pérez-Hugalde, C., Picard, A., & Chaya, C. (2016). Influence of first position effect on emotional response. *Food Quality and Preference, 49*, 189–196. https://doi.org/10.1016/j.foodqual.2015.12.009

Dorfman, D. D., & Alf, E (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—Rating-method data. *Journal of Mathematical Psychology, 6*(3), 487–496. https://doi.org/10.1016/0022-2496(69)90019-4

Endrizzi, I., Gasperi, F., Rødbotten, M., & Næs, T. (2014). Interpretation, validation and segmentation of preference mapping models. *Food Quality and Preference, 32*, 198–209. https://doi.org/10.1016/j.foodqual.2013.10.002

Ennis, D. M., Boelens, H., Haring, H., & Bowman, P. (1982). Multivariate analysis in sensory evaluation. *Food Technology, 36*(11), 83–90.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Gotow, N., Nagai, Y., Taguchi, T., Kino, Y., Ogino, H., & Kobayakawa, T. (2023). Change in preference for vegetables and their perceptual attributes as a function of age and pickiness. *Food Research International, 170*, Article 112967. https://doi.org/10.1016/j.foodres.2023.112967

Goyat, S. (2011). The basis of market segmentation: A critical review of literature. *European Journal of Business and Management, 3*(9), 45–54.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. 1969–2012). New York: Wiley.

Grunert, K. G., Bech-Larsen, T., & Bredahl, L. (2000). Three issues in consumer quality perception and acceptance of dairy products. *International Dairy Journal, 10*(8), 575–584. https://doi.org/10.1016/S0958-6946(00)00085-6

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide* (3rd ed.). Routledge. https://doi.org/10.4324/9781003203636

Hayes, R. J., & Simko, I. (2016). Breeding lettuce for improved fresh-cut processing. *Acta Horticulturae, 1141*, 65–76. https://doi.org/10.17660/ActaHortic.2016.1141.7

Hildebrandt, G. H., Dominguez, B. L., Schork, M. A., & Loesche, W. J. (1997). Functional units, chewing, swallowing, and food avoidance among the elderly. *The Journal of Prosthetic Dentistry, 77*(6), 588–595. https://doi.org/10.1016/S0022-3913(97)70100-8

Hottenstein, A. W., Taylor, R., & Carr, B. T. (2008). Preference segments: A deeper understanding of consumer acceptance or a serving order effect? *Food Quality and Preference, 19*(8), 711–718. https://doi.org/10.1016/j.foodqual.2008.04.004

Jaeger, S. R., Rossiter, K. L., Wismer, W. V., & Harker, F. R. (2003). Consumer-driven product development in the kiwifruit industry. *Food Quality and Preference, 14*(3), 187–198. https://doi.org/10.1016/S0950-3293(02)00053-8

Jeong, Y.-N., Kang, B.-A., Jeong, M.-J., Song, M.-J., Hautus, M.-J., & Lee, H.-S. (2016). Sensory discrimination by consumers of multiple stimuli from a reference: Stimulus configuration in A-not AR and constant-ref. duo-trio superior to triangle and unspecified tetrad? *Food Quality and Preference, 47*, 10–22. https://doi.org/10.1016/j.foodqual.2015.06.021

Kim, I.-A., Hopkinson, A., van Hout, D., & Lee, H.-S. (2017a). A novel two-step rating-based 'double-faced applicability' test. Part 1: Its performance in sample discrimination in comparison to simple one-step applicability rating. *Food Quality and Preference, 56*, 189–200. https://doi.org/10.1016/j.foodqual.2016.10.010

Kim, I.-A., Hopkinson, A., van Hout, D., & Lee, H.-S. (2017b). A novel two-step rating-based 'double-faced applicability'test. Part 2: Introducing a novel measure of affect magnitude (d' A) for profiling consumers' product usage experience based on signal detection theory. *Food Quality and Preference, 59*, 141–149. https://doi.org/10.1016/j.foodqual.2017.02.011

Kim, I.-A., van Hout, D., & Lee, H.-S. (2018a). Signal detection-based satisfaction measure of the holistic product usage experience with and without the 'double-faced applicability' test. *Food Quality and Preference, 68*, 40–49. https://doi.org/10.1016/j.foodqual.2018.02.005

Kim, J.-N., Yun, B.-H., Lee, Y.-J., Jang, Y.-A., & Lee, H.-S. (2026). Integrated sensory profiling of fresh lettuce varieties: Combining affect magnitude profiling with generalized descriptive analysis. *Food Quality and Preference, 105694*. https://doi.org/10.1016/j.foodqual.2025.105694

Kim, K., Chun, I.-J., Suh, J.-H., & Sung, J. (2023). Relationships between sensory properties and metabolomic profiles of different apple cultivars. *Food Chemistry: X, 18*, Article 100641. https://doi.org/10.1016/j.fochx.2023.100641

Kim, M.-A., Dessirier, J. M., van Hout, D., & Lee, H.-S. (2015). Consumer context-specific sensory acceptance tests: Effects of a cognitive warm-up on affective product discrimination. *Food Quality and Preference, 41*, 163–171. https://doi.org/10.1016/j.foodqual.2014.11.019

Kim, M.-A., & Lee, H.-S. (2015). Duo-trio difference-preference test with two replications: Use of psychological biases for measuring meaningful preference. *Journal of Sensory Studies, 30*(3), 211–224. https://doi.org/10.1111/joss.12152

Kim, M.-A., Lee, Y.-J., Kim, M.-S., & Lee, H.-S. (2022). Reminder–preference test, affective difference-preference test using reference framing with a brand: 1. Sensitivity comparisons with the same–different difference–preference test. *Food Research International, 155*, Article 111065. https://doi.org/10.1016/j.foodres.2022.111065

Kim, M.-A., Sim, H.-M., & Lee, H.-S. (2015). Affective discrimination methodology: Determination and use of a consumer-relevant sensory difference for food quality maintenance. *Food Research International, 70*, 47–54.

Kim, M.-A., van Hout, D., Dessirier, J. M., & Lee, H.-S. (2018b). Degree of satisfaction-difference (DOSD) method for measuring consumer acceptance: A signal detection measurement with higher reliability than hedonic scaling. *Food Quality and Preference, 63*, 28–37. https://doi.org/10.1016/j.foodqual.2017.07.012

Kim, M.-A., van Hout, D., & Lee, H.-S. (2018c). Degree of satisfaction-difference (DOSD) method for measuring consumer acceptance: Comparative and absolute measures of satisfaction based on signal detection theory. *Food Quality and Preference, 68*, 167–172. https://doi.org/10.1016/j.foodqual.025

Kim, M.-A., van Hout, D., Zandstra, E. H., & Lee, H.-S. (2019). Consumer acceptance measurement focusing on a specified sensory attribute of products: Can the attribute-specified degree of satisfaction-difference (DOSD) method replace hedonic scaling? *Food Quality and Preference, 75*, 198–208. https://doi.org/10.1016/j.foodqual.2019.03.009

Lawless, H. T., & Heymann, H. (2010). Sensory evaluation of food: Principles and practices. *Springer Science & Business Media.*. https://doi.org/10.1007/978-1-4419-7452-5

Lee, H.-S., & van Hout, D. (2009). Quantification of sensory and food quality: The R-index analysis. *Journal of Food Science, 74*(6), R57–R64. https://doi.org/10.1111/j.1750-3841.2009.01204.x

Lee, H.-S., van Hout, D., & O'mahony, M.. (2007). Sensory difference tests for margarine: A comparison of R-indices derived from ranking and A-not A methods considering response bias and cognitive strategies. *Food Quality and Preference, 18*(4), 675–680. https://doi.org/10.1016/j.foodqual.2006.10.003

Lee, Y.-J., Kim, I.-A., van Hout, D., & Lee, H.-S. (2021). Investigating effects of cognitively evoked situational context on consumer expectations and subsequent consumer satisfaction and sensory evaluation. *Food Quality and Preference, 94*, Article 104330. https://doi.org/10.1016/j.foodqual.2021.104330

Lee, Y.-J., Lim, H.-J., & Lee, H.-S. (2025). Complementary evaluation approaches for sensory acceptance: Monadic liking vs. paired satisfaction in relation to bias sensitivity, data outputs, and consumer segmentation. *Food Quality and Preference, 105680*. https://doi.org/10.1016/j.foodqual.2025.105680

Lee, Y.-S., & Meullenet, J. F. (2010). Comparison of eliminating first order samples for minimizing first serving order bias to data corrections. *Food Science and Biotechnology, 19*, 703–709. https://doi.org/10.1007/s10068-010-0099-6

Lichters, M., Möslein, R., Sarstedt, M., & Scharf, A. (2021). Segmenting consumers based on sensory acceptance tests in sensory labs, immersive environments, and natural consumption settings. *Food Quality and Preference, 89*, Article 104138. https://doi.org/10.1016/j.foodqual.2020.104138

Ling, M., Bai, X., Cui, D., Shi, Y., Duan, C., & Lan, Y. (2023). An efficient methodology for modeling to predict wine aroma expression based on quantitative data of volatile compounds: A case study of oak barrel-aged red wines. *Food Research International, 164*, Article 112440. https://doi.org/10.1016/j.foodres.2022.112440

Løkke, M. M., Seefeldt, H. F., & Edelenbos, M. (2012). Freshness and sensory quality of packaged wild rocket. *Postharvest Biology and Technology, 73*, 99–106. https://doi.org/10.1016/j.postharvbio.2012.06.004

MacFie, H. J. H. (2016, March). *Sensometrics has got to move on. In 13th Sensometrics conference, Brighton, UK, July (pp. 26-29)*.

MacFie, H. J. H., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies, 4*(2), 129–148. https://doi.org/10.1111/j.1745-459X.1989.tb00463.x

Marsh, K., Rossiter, K., Lau, K., Walker, S., Gunson, A., & MacRae, E. A. (2003). *The use of fruit pulps to explore flavour in kiwifruit*. https://doi.org/10.17660/ActaHortic.2003.610.30

Mazur, J., Drabek, R., & Goldman, A. (2018). Hedonic contrast effects in multi-product food evaluations differing in complexity. *Food Quality and Preference, 63*, 159–167. https://doi.org/10.1016/j.foodqual.2017.06.016

Meilgaard, M. C., Carr, B. T., & Civille, G. V. (1999). *Sensory evaluation techniques*. CRC press. https://doi.org/10.1201/9781003040729

Moskowitz, H. R., Beckley, J. H., & Resurreccion, A. V. (2012). *Sensory and consumer research in food product design and development. Chapter 9* (2nd ed.). Ames, Iowa: Blackwell Publishing.

Müller, H., & Hamm, U. (2014). Stability of market segmentation with cluster analysis – A methodological approach. *Food Quality and Preference, 34*, 70–78. https://doi.org/10.1016/j.foodqual.2013.12.004

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification, 31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

O'Brien, R., Simon, J. E., & Tepper, B. J. (2022). How do trained panelists characterize baby leafy greens? A comparison of descriptive analysis and napping. *Journal of Food Science, 87*(1), 396–414. https://doi.org/10.1111/1750-3841.15985

Onwezen, M. C., & Bartels, J. (2011). Which perceived characteristics make product innovations appealing to the consumer? A study on the acceptance of fruit innovations using cross-cultural consumer segmentation. *Appetite, 57*(1), 50–58. https://doi.org/10.1016/j.appet.2011.03.011

Onwezen, M. C., Reinders, M. J., van der Lans, I. A., Sijtsema, S. J., Jasiulewicz, A., Guardia, M. D., & Guerrero, L. (2012). A cross-national consumer segmentation based on food benefits: The link with consumption situations and food perceptions. *Food Quality and Preference, 24*(2), 276–286. https://doi.org/10.1016/j.foodqual.2011.11.002

Poelman, A. A. M., & Delahunty, C. M. (2011). The effect of preparation method and typicality of colour on children's acceptance for vegetables. *Food Quality and Preference, 22*(4), 355–364. https://doi.org/10.1016/j.foodqual.2011.01.001

Schilling, M. W., & Coggins, P. C. (2007). Utilization of agglomerative hierarchical clustering in the analysis of hedonic scaled consumer acceptability data. *Journal of Sensory Studies, 22*(4), 477–491. https://doi.org/10.1111/j.1745-459X.2007.00121.x

Schröder, M. J. (2003). Food quality and consumer value: Delivering food that satisfies. *Springer Science & Business Media*. https://doi.org/10.1007/978-3-662-07283-7

Serrano-Megías, M., & López-Nicolás, J. M. (2006). Application of agglomerative hierarchical clustering to identify consumer tomato preferences: Influence of physicochemical and sensory characteristics on consumer response. *Journal of the Science of Food and Agriculture, 86*(4), 493–499. https://doi.org/10.1002/jsfa.2392

Shewfelt, R. L. (2014). Measuring quality and maturity. In *Postharvest handling* (pp. 387–410). Academic Press. https://doi.org/10.1016/B978-0-12-408137-6.00014-4.

Shu, N., Chen, X., Sun, X., Cao, X., Liu, Y., & Xu, Y. J. (2023). Metabolomics identify landscape of food sensory properties. *Critical Reviews in Food Science and Nutrition, 63*(27), 8478–8488. https://doi.org/10.1080/10408398.2022.2062698

Swegarden, H., Stelick, A., Dando, R., & Griffiths, P. D. (2019). Bridging sensory evaluation and consumer research for strategic leafy Brassica (Brassica oleracea) improvement. *Journal of Food Science, 84*(12), 3746–3762. https://doi.org/10.1111/1750-3841.14831

Talavera-Bianchi, M., Chambers, E, IV, & Chambers, D. H. (2010). Lexicon to describe flavor of fresh leafy vegetables. *Journal of Sensory Studies, 25*(2), 163–183. https://doi.org/10.1111/j.1745-459X.2009.00249.x

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning, 20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Ulrich, D., & Olbricht, K. (2016). A search for the ideal flavor of strawberry-comparison of consumer acceptance and metabolite patterns in Fragaria× ananassa Duch. *Journal of Applied Botany and Food Quality, 89*. https://doi.org/10.5073/JABFQ.2016.089.029

Varela, P. (2013). 11 application of multivariate statistical methods during new product development–case study: Application of principal component analysis and hierarchical cluster analysis on consumer liking data of orange juices. *Mathematical and statistical methods in food science and technology, 187*. https://doi.org/10.1002/9781118434635.ch11

Vigneau, E., & Chen, M. (2014). ClustVarLV: Clustering of variables around Latent Variables. http://CRAN.R-project.org/package=ClustVarLV. R package version 1.3.1.

Vigneau, E., Chen, M., & Qannari, E. M. (2015). ClustVarLV: An R package for the clustering of variables around latent variables. *The R Journal, 7*(2), 134. https://doi.org/10.32614/RJ-2015-026

Vigneau, E., & Qannari, E. M. (2002). Segmentation of consumers taking account of external data. A clustering of variables approach. *Food Quality and Preference, 13*(7–8), 515–521. https://doi.org/10.1016/S0950-3293(02)00048-4

Vigneau, E., & Qannari, E. M. (2003). Clustering of variables around latent components. *Communications in Statistics-Simulation and Computation, 32*(4), 1131–1150. https://doi.org/10.1081/SAC-120023882

Vigneau, E., Qannari, E. M., Navez, B., & Cottet, V. (2016). Segmentation of consumers in preference studies while setting aside atypical or irrelevant consumers. *Food Quality and Preference, 47*, 54–63. https://doi.org/10.1016/j.foodqual.2015.02.008

Wajrock, S., Antille, N., Rytz, A., Pineau, N., & Hager, C. (2008). Partitioning methods outperform hierarchical methods for clustering consumers in preference mapping. *Food Quality and Preference, 19*(7), 662–669. https://doi.org/10.1016/j.foodqual.2008.06.002

Wakeling, I. N., & MacFie, H. J. H. (1995). Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of k samples from t may be tested. *Food Quality and Preference, 6*(4), 299–308. https://doi.org/10.1016/0950-3293(95)00032-1

Westad, F., Hersleth, M., & Lea, P. (2004). Strategies for consumer segmentation with applications on preference data. *Food Quality and Preference, 15*(7–8), 681–687. https://doi.org/10.1016/j.foodqual.2004.04.004

Xie, X., Tang, T., Wang, W., Tang, X., Zhang, J., & Wang, Z. (2021). Metabolomics clarify the compounds contributing to the quality of apples among different regions in China. *Journal of Food Processing and Preservation, 45*(1), Article e15054. https://doi.org/10.1111/jfpp.15054

Yenket, R., Chambers, E., & Johnson, D. E. (2011). Statistical package clustering may not be best for grouping consumers to understand their most liked products. *Journal of Sensory Studies, 26*(3), 209–225. https://doi.org/10.1111/j.1745-459X.2011.00337.x

Yoon, H.-J. (2022). *Effect of a pre-session of tasting samples on results of food choice experiments with eye-tracking using RTD coffee packages [master's thesis, Ewha Womans University]*. Ewha Womans University Library.

Yun, B.-H. (2023). *Investigation of consumer acceptability of ten cultivars of lettuce (Lactuca sativa L.) using the 9-point hedonic scale and the degree of satisfaction-difference method [master's thesis, Ewha Womans University]. Ewha Womans University library*.