Quantifying Feel, Fragrance And Finish: A Review of Sensory Threshold In Cosmetics

Suryaneta[1*], Kiki Fibrianto[2]

[1]Cosmetic Engineering Bachelor Program, Institut Teknologi Sumatera, South Lampung, 35365, Indonesia
[2]Jurusan Teknologi Hasil Pertanian, Fakultas Teknologi Pertanian, Universitas Brawijaya Jl. Veteran, Malang 65145, Indonesia
*Corresponding Author:
Email: suryaneta@km.itera.ac.id

Abstract.

Sensory experience determines whether cosmetic products are adopted, repurchased, and trusted. Yet the knowledge needed to set defensible sensory targets is scattered across psychophysics, descriptive analysis, and category-specific practice. This review consolidates the concept of "thresholds" for cosmetics and explains how to use them to guide formulation, quality control, and claims. Author define detection and recognition thresholds as performance-based points on a psychometric function that ensure key notes and tactile cues are truly perceivable and correctly identified. Author then describe the difference threshold, or just noticeable difference, as the smallest reliable change from a reference and show how JNDs translate directly into specification bands that control batch-to-batch drift. Because perceptibility does not guarantee liking, we integrate consumer-facing acceptance and rejection thresholds to locate intensity regions that preserve preference and avoid penalties in market. Methodologically, the review emphasizes bias-resistant forced-choice designs, supported by adjustment and categorical procedures, and shows how threshold estimation aligns with descriptive sensory programs already used for creams and lotions across realistic stages of use. Taken together, these tools provide a practical bridge from small compositional or process changes to user-relevant discriminability and acceptance. Author conclude with priorities for practice, including disciplined panel management, tighter linkage between laboratory thresholds and in-use temporal profiles, and opportunities for mobile or at-home protocols that capture real-world experiences.

Keywords: *Sensory threshold; detection threshold; recognition threshold; just noticeable difference and cosmetic product development.*

I. INTRODUCTION

Sensory experience is the currency of the cosmetic industry (Keightley et al., 2016). The way a product looks, smells, and feels determines whether it is applied as intended, whether it earns a place in a daily routine, and whether it is bought again (Theofanides & Kerasidou, 2012). To manage these outcomes, developers need more than descriptive language; they need quantitative limits that tell them when a signal becomes perceivable, when its quality is identifiable, how much it must change to be noticed, and where liking begins to decline (Ardoin et al., 2020). These limits are sensory thresholds, and they provide the bridge from psychophysics to day-to-day formulation, process control, and claim support (Drake et al., 2023). In a psychophysical frame, a threshold is not a vague impression but a reproducible point on a performance curve. Detection thresholds locate the onset of perceivability. Recognition thresholds ensure that a note, texture, or color is correctly identified rather than merely sensed. Difference thresholds, or just noticeable differences, quantify the smallest reliable change from a reference and translate directly into specification bands that guard against batch-to-batch drift. Because being noticeable does not guarantee being liked, acceptance and rejection thresholds complete the picture by defining the intensity region that preserves preference and the point at which consumers turn away (Lawless & Heymann, 2010). Despite the centrality of these concepts, guidance for cosmetics remains scattered across diverse sources.

Many standards and textbooks teach how to profile attributes with trained panels, yet few consolidate the definitions, boundary conditions, and measurement choices needed to set defensible thresholds for fragrance, texture, color, and visual finish (Pensé-Lhéritier, 2015). As a result, teams often struggle to convert small compositional or process adjustments into consumer-relevant decisions, or to align instrumental targets with what users can actually perceive during pick-up, rub-out, application, and after-feel (Guest et al., 2013).This review addresses that gap. Author bring together threshold definitions that are grounded in observable performance, explain their implications for topical products, and describe the methods most suitable for cosmetics, with emphasis on bias-resistant forced-choice designs supported by adjustment and categorical procedures. Author show how threshold estimation can be embedded within existing descriptive programs for creams and lotions so that numbers derived in the laboratory become actionable targets for formulation and quality control. Author also connect perceptual limits to consumer acceptance, highlighting how rejection and acceptance thresholds provide practical guardrails for brand signatures over shelf life and across manufacturing lots.The article follows a traditional literature review approach. It synthesizes evidence from psychophysics, sensory analysis, and category-specific practice to produce a single, cosmetics-ready reference. The intended contribution is clarity: what each threshold means, how to measure it well, and how to use those measurements to make better product decisions that are both perceptible to users and acceptable in market.

Review Approach

This article follows a traditional literature review. Author synthesize definitions of detection, recognition, and difference thresholds within a psychophysical framework; summarize core measurement procedures such as forced choice, adjustment, and categorical classification; and integrate these with descriptive standards that are already used in cosmetics. Our scope explicitly includes consumer-facing acceptance and rejection limits so that perceptual boundaries are connected to hedonic consequences. The goal is a single, cosmetics-ready reference that clarifies what each threshold means, how to measure it with bias-resistant designs, and how to use the resulting numbers to guide formulation, quality control, and claims.

What A “Threshold” Means in Cosmetic Context

Across sensory science, a threshold is a quantitative boundary on an intensity continuum that marks a reproducible shift in perceptual performance. In experimental practice, this boundary is tied to a criterion point on a psychometric function so that the estimate rests on observable responses rather than introspective reports. For cosmetics, this framing allows developers to connect stimulus intensity—fragrance load, color depth, viscosity, or surface slip—to the probability that users will detect, recognize, or discriminate what they feel, see, or smell during real use. detection

Types of Thresholds and How They Are Defined

Detection Threshold

The detection threshold is the minimal intensity of a stimulus that can be reliably discerned as present rather than absent when evaluated with a formal psychophysical task (Bi & Ennis, 1998; International Organization for Standardization, 2012) . In forced choice paradigms, this threshold is defined with respect to a criterion point on the psychometric function, typically the level that yields a probability of detection that is one half above chance once the chance level has been removed from the observed proportion correct (Klein, 2001). This definition anchors the threshold to observer performance rather than subjective report and is widely adopted in standard practice for sensory threshold work (Macmillan & Creelman, 2004). Measurement commonly uses an n alternative forced choice design with an ascending concentration series method of limits. A frequent choice is the three alternative forced choice format in which one sample contains the target stimulus and two are controls (Meilgaard, 1991). The series is constructed over at least five concentration steps, often six or seven, with each step differing by a factor of two to four (El Ouyoun Najm et al., 2010). Replication at each step is recommended in order to estimate the psychometric function with sufficient precision, after which the threshold can be obtained graphically or by regression using probit or related transforms. These principles are codified in American Society for Testing and Materials practices for odor and taste threshold determination and for calculation of individual and group thresholds.

Forced choice procedures are preferred because they reduce response bias and have superior reliability compared with yes or no judgments. Classic experimental analyses show that thresholds derived from forced choice data are more stable across sessions and less vulnerable to irrelevant variations in procedure. This reliability advantage supports their use when robust estimates are required for formulation targets and specification limits (Macmillan & Creelman, 2004). Quality of estimation depends on panel design and execution. Increasing the number of observers and the number of stimulus presentations improves the confidence of the threshold estimate, while panel composition and prior exposure to the stimulus class can shift the measured values. Methodological guidance therefore emphasizes adequate sample sizes, replication, and assessor training for the target modality.Although the detection threshold concept is general, its implementation is modality specific. For taste, international standards describe how to prepare basic taste series and how to determine detection and recognition limits under controlled conditions. For odor and fragrance, forced choice ascending concentration series methods are recommended to locate the lowest concentration at which a note is detectable. These same principles extend to visual and tactile cues relevant to cosmetics, where the detection threshold would correspond to the faintest color that departs from a clear base or the weakest tactile cue that can be sensed on skin during product application.

Recognition Threshold

The recognition threshold is the minimal stimulus intensity at which a person can correctly identify the quality or category of a sensation rather than merely detect its presence (Hohl et al., 2014). In formal sensory practice this construct is anchored to identification tasks and is paired with the detection threshold in standard methods for assessing sensory acuity (Webb et al., 2015). Guidance for taste sensitivity specifies procedures that yield both detection and recognition thresholds, and notes that for some individuals the two may coincide at the same concentration level (Trius-Soler et al., 2020). Recognition is operationalized with identification paradigms in which one of several possible signals is presented and the observer indicates which quality is present. Psychophysical treatments show that multi alternative identification produces efficient estimates because chance performance is low and the psychometric function can be fitted with adequate precision for threshold inference (García-Pérez et al., 2011).Modality specific protocols illustrate how recognition thresholds are obtained in practice. For olfaction, odor recognition and odor matching exercises are widely used for screening and training, and the University of Pennsylvania Smell Identification Test provides a forced choice identification framework that quantifies smell function across defined levels. Age related differences are incorporated in the evaluation, which underscores the importance of demographic factors when interpreting recognition limits. For taste, the international standard for taste sensitivity details preparation of basic taste series and the determination of both detection and recognition thresholds under controlled conditions.

Factors that shift recognition thresholds include physiological state, adaptation, prior exposure, and training. Reviews of assessor management highlight that many studies evaluate gustatory and olfactory function using detection or recognition thresholds, and they recommend control of timing and recent intake to minimize variability. Adaptation and build up effects are particularly salient for odor and taste and can bias identification unless rest and rinsing protocols are enforced.The construct is directly relevant to cosmetics where a brand signature note, a textural cue, or a color quality must be identifiable under intended use. Assessor screening for cosmetic and other nonfood products routinely includes aroma recognition exercises with reference compounds specified in international guidance, which ensures that panelists can identify defined targets such as benzaldehyde for almond or vanilla for vanilla. This linkage between recognition performance and reference standards supports robust estimation of recognition thresholds for fragrance and other modalities.Therefore, the recognition threshold formalizes the point on the intensity continuum at which quality identification becomes reliable. It is measured with identification tasks that can be analyzed through a psychometric function, it is sensitive to methodological and physiological influences, and it provides actionable limits for ensuring that desired notes, textures, or appearance cues are truly recognized in cosmetic products.

Difference Threshold or Just Noticeable Difference (JND)

The just noticeable difference is the smallest change in a stimulus that yields a reliable change in discrimination performance and is therefore interpreted as a minimal perceptible difference. In classical psychophysics, the JND is defined at a criterion point on the psychometric function, commonly the stimulus disparity that produces correct discrimination on one half of the trials once chance performance is taken into account (Bausenhart et al., 2012). This operational definition links the construct to observable performance rather than introspective report (Macmillan & Creelman, 2004). Across many modalities, the magnitude of the JND scales with the reference intensity such that a roughly constant proportion of the base level is required to produce a noticeable difference, a regularity summarized by Weber's law and often used to express a Weber fraction for the attribute of interest (Deco et al., 2007). This proportional rule provides a compact way to anticipate how much the texture firmness, viscosity, color depth, or fragrance concentration must change before consumers can perceive a difference under controlled conditions. JNDs are estimated with formal discrimination tasks that fit a psychometric function. Methods include constant stimuli with two interval or three alternative forced choices, and adaptive sequences, each chosen to control response bias and provide efficient estimation.

Standards and guidance for threshold work specify forced choice designs with ascending concentration series, together with calculation procedures for individual and group thresholds, which can be adapted from taste and odor contexts to cosmetic attributes such as fragrance or low contrast color differences. In practice, recognition that the JND depends on decision processes as well as sensory input has motivated the use of signal detection theory to separate sensitivity from response bias when fitting discrimination data. This framework clarifies why the observed JND can shift with motivation, attention, or task structure even when the underlying sensory capacity is unchanged, and it supports principled comparisons across panels and procedures. For cosmetics, JNDs translate directly into formulation and quality control targets because they define the minimal physical change that users can perceive as a difference between otherwise similar products. Descriptive guides for skin creams and lotions provide structured attribute vocabularies and time of use staging, which can be combined with JND based difference testing to judge whether a change in rheology modifier, emollient level, or fragrance load is perceptible during pick up, rub out, or after feel. This integration links small instrumental or compositional shifts to consumer discriminability in a category specific way. Finally, using JNDs to set specification bands is valuable throughout development and scale up. If batch to batch variability remains below the relevant JND for a key attribute, perceptible differences are unlikely in blinded comparison, whereas excursions above the JND indicate a risk that consumers will notice and possibly prefer one lot over another. Embedding JND estimation within standardized threshold protocols therefore provides actionable limits that support robust sensory performance in market.

Acceptance and Rejection Thresholds

Acceptance and rejection thresholds define consumer facing limits on stimulus intensity at which overall liking begins to decline and at which rejection becomes evident (Lima Filho et al., 2015). These limits are estimated from consumer responses and therefore complement perceptual thresholds by indicating where changes in fragrance, texture, color, or appearance begin to compromise product acceptance and where rejection emerges (Filho et al., 2018).The consumer rejection threshold was introduced to locate the stimulus level at which preference shifts away from a control, using paired preference tests within a constant stimulus framework and testing a series of increasing or decreasing intensities against a fixed standard (Filho et al., 2017). This procedure formalizes the point at which dislike begins to dominate choice behavior. Subsequent work has argued that acceptance tests using a hedonic scale are more appropriate when the goal is to identify the onset of acceptance loss or outright rejection, because lower preference does not necessarily imply sensory rejection (Gamba et al., 2021). This reasoning led to the development of two complementary measures obtained directly from acceptance data, namely the compromised acceptance threshold that marks the initial impairment of liking and the rejection threshold that marks transition from acceptance to rejection. Methodologically, these thresholds are determined by presenting consumers with a control product and with samples that vary the target stimulus across a defined range, then modelling the acceptance response to estimate the cross over points.

Compared with preference based rejection thresholds, acceptance based estimates provide greater reliability for identifying the points at which acceptance begins to decline and at which rejection is established.Acceptance and rejection thresholds do not coincide with detection limits. A product may be perceptibly different without any impact on liking, and rejection often requires intensities that exceed the lower detection boundary. This distinction explains why detection methods are insufficient for guiding consumer acceptability limits during formulation. These consumer thresholds have broad practical value. They support quality control by defining tolerable bands around a target, inform reformulation by identifying safe ranges for key sensory drivers, and assist shelf life studies by indicating the stimulus drift that will jeopardize acceptance. Their proposed use extends beyond foods to cosmetic and pharmaceutical categories, where they enable evidence based setting of fragrance load, viscosity cues, and color depth.Finally, acceptance and rejection limits are not universal constants. Individual differences in hedonic response and in the shape of the liking curve create segments that prefer different intensity levels for the same attribute, which implies that thresholds and specification bands may need to be tailored for distinct consumer groups or market positions.

How Thresholds Relate to Cosmetic Products and Real Use

Mapping thresholds to stages of use

Standards for skin creams and lotions break the experience into delivery, pick-up, rub-out, application, and after-feel, and provide assessor selection, training, and scaling guidance. Positioning detection, recognition, JND, and acceptance limits within these stages makes the numbers actionable: for example, ensuring detectability of a top-note at first impression, recognizability of a powdery cue during rub-out, or keeping after-feel residue below the acceptance boundary.

Texture and After-Feel

Attributes such as wetness, spreadability, thickness, oiliness, greasiness, absorbency, gloss, slipperiness, and residue are specified and tracked by trained assessors using category vocabularies. Threshold thinking connects these descriptive measures to consumer-relevant limits. A lotion’s capacity to absorb without leaving undesirable residue is a clear example: descriptive protocols quantify how quickly a product integrates into skin and how much residue remains, while acceptance thresholds indicate when residue intensity begins to erode liking.

Fragrance and Color Signatures

Recognition thresholds help preserve identifiable yet balanced brand signatures in fragrance or finish, while JNDs guide the smallest meaningful change in color depth or accord strength across reformulations. Acceptance and rejection limits then set the guardrails that prevent over-intense notes or overly opaque finishes from triggering penalties in liking. Together, these thresholds support consistent sensory branding across batches and over shelf life.

From Numbers to Specifications

Forced-choice thresholds provide bias-resistant lower bounds for perceivability; JNDs translate micro-changes into consumer-relevant discriminability; acceptance and rejection thresholds protect liking. Embedding this trio within established descriptive programs yields defensible targets for formulation, quality control, and change-control decisions in creams, lotions, tinted finishes, and other topical formats.

II. METHODS

Psychophysical Approaches

Psychophysical approaches estimate sensory thresholds by linking observer performance to controlled changes in stimulus intensity and by fitting a psychometric function that defines the criterion point taken as the threshold. In practice, three families of procedures dominate threshold work in cosmetics and related domains, namely forced choice, adjustment, and classification. Authoritative standards and textbooks describe how these procedures minimize bias, deliver reproducible estimates, and translate directly into formulation targets. Forced choice methods present the observer with a small set of alternatives on each trial and require a choice of the stimulus that contains the signal. A common implementation for odor or taste is the three alternative forced choice ascending series described in ASTM E679, which specifies sample preparation, replication at each concentration step, and calculation of individual and group thresholds for either detection or recognition. The ascending sequence is continued until performance reaches the prespecified criterion above chance, after which the threshold is estimated from the fitted psychometric function. The chief advantage of forced choice is resistance to response bias, a property that is formalized in detection theory and that improves the stability of estimates across sessions and laboratories.

Adjustment methods allow the observer to vary the stimulus continuously or in small steps until the specified criterion is met, for example the point at which a faint fragrance becomes just detectable or a textural cue becomes just identifiable. Modern psychophysics texts describe adjustment as efficient for preliminary mapping and for device calibration, while emphasizing the need for multiple reversals, careful instruction, and independent replications to control expectancy and adaptation. When the adjustment track is recorded, a threshold can be derived from the mean of the final settings or from a staircase convergence rule. These procedures are widely used as quick checks before more rigorous forced choice testing (ASTM, 2015). Classification methods convert perceived intensity into ordered categories and thereby locate threshold regions on a categorical scale. Category scaling and related labeled magnitude procedures are frequently employed to anchor the lower end of sensation to a practical reporting scheme, especially when training assessors or when relating instrumental readings to perceived strength. Although classification does not produce a threshold in the same inferential sense as forced choice, it provides valuable boundary information for detectability and recognition and can be combined with acceptance diagnostics during development.

Reviews and method chapters in sensory analysis outline the design of category tasks, the statistical treatment of category data, and the use of labeled magnitude scales for cross modality comparison. Across these approaches, good practice requires control of context, temperature, and presentation sequence, along with assessor familiarization. ISO 3972 offers detailed instructions for taste sensitivity that are directly transferable to other modalities, including explicit definitions of detection, recognition, and difference thresholds and recipes for test solutions, randomization, and assessor management. These controls reduce variability due to physiological state, adaptation, or learning, and they increase the validity of threshold estimates used in formulation and quality control. In summary, forced choice provides bias resistant and statistically tractable detection and recognition thresholds, adjustment supplies efficient preliminary estimates and device calibration within well controlled tracks, and classification yields categorical boundaries that are useful for training, scale alignment, and translation to consumer language. Together, these psychophysical procedures offer a coherent toolkit for establishing quantitative sensory targets for fragrance, texture, color, and appearance in cosmetic product development.

Descriptive Sensory Analysis

Descriptive sensory analysis provides an analytic framework to characterize appearance, fragrance, texture, and related attributes with trained human assessors who function as measurement instruments. Classical and modern references position this family of methods as objective and repeatable, with broad utility for quality control, product comparison, understanding consumer response, sensory mapping for opportunity identification, and claims substantiation. Quantitative Descriptive Analysis characterizes products by developing a category specific vocabulary in the language of consumers and by training screened likers and users to rate intensities on structured scales. Method texts describe two core advantages that are particularly relevant for cosmetics. First, the vocabulary is expressed in everyday terms that are easily understood across technical and marketing teams, which facilitates translation of sensory findings into commercial direction. Second, panel formation and language development can be accomplished in a relatively short timeline, which supports iterative development cycles. These properties make it straightforward to relate QDA attribute scores to hedonic data and to preference mapping in order to target consumer relevant intensity levels.

The Spectrum method represents a technical expert approach that emphasizes standardized terminology, fixed intensity references, and tight panel calibration so that assessors behave as a consistent measuring device. Authoritative chapters stress the central role of a skilled panel leader who understands the product category, teaches protocol and scaling, and provides continuous feedback to refine both language and scaling precision. The same sources recommend ongoing coaching and performance review to maintain sensitivity to small qualitative and quantitative differences.Category specific guidance for skin care products integrates both approaches within a single standard. The ASTM guide for creams and lotions describes procedures to define attributes and to measure their intensities and changes over time, then links those results to formulation direction, ingredient substitution, competitive assessment, research guidance, and claim substantiation. The same document provides criteria for selecting and training assessors and describes evaluation of the product alone and on skin. It also defines stages of use such as pick up, rub out, delivery, and application, and recommends the use of reference anchors for scale alignment across sessions.

In practice, descriptive programs for cosmetics follow a structured panel pathway that includes recruitment and pre-screening, screening and selection, attribute generation, standardization of perception, rating method familiarization, development of rating skills, competence checking, and formal product assessment. Adhering to this progression improves reliability and enables consistent comparison across prototypes and production lots. These methodological principles have been demonstrated in cosmetic categories beyond skin care. Lexicon development for lip products illustrates how trained panels describe appearance and texture during and after use, and how the resulting language connects to existing skinfeel guidance for creams and lotions. This cross-category coherence supports a unified descriptive vocabulary that can be mapped to consumer language in QDA and to calibrated references in Spectrum panels. Together, QDA and Spectrum constitute complementary descriptive toolkits for measuring sensory attributes that underpin thresholds in cosmetics. QDA provides rapid, consumer understandable profiling that is easily linked to acceptance models, while Spectrum supplies calibrated, reference anchored measurement that strengthens reproducibility and cross study comparability. Using the ASTM skin care guide as a procedural backbone ensures that texture, fragrance, and color are characterized across defined stages of use with trained assessors and defensible intensity scales.

Dynamic Sensory Profiling

Dynamic sensory profiling quantifies how multiple sensations unfold over the course of a product experience. Two families of temporal methods are commonly employed. Time intensity traces the evolution of a single attribute, whereas Temporal Dominance of Sensations monitors several attributes and records which sensation is momentarily dominant as time progresses. These approaches capture changes that occur from the first visual or olfactory impression through application and into the after feel period, which is especially relevant for creams, lotions, and other topical formats. Temporal Dominance of Sensations is implemented by asking assessors to select, at each instant, the one attribute that best describes the dominant sensation from a predefined list. The result is a dominance curve for each attribute and a dominance rate that can be analyzed over the full evaluation window. Contemporary cosmetic studies apply explicit interpretive rules, for example a minimum of ten percent of the total evaluation time, in order to decide whether a temporal signal is meaningful. This rule helps analysts distinguish sustained dominance from brief transients and supports reliable product comparisons. Design of a Temporal Dominance of Sensations protocol requires careful definition of the attribute list and alignment of the evaluation timeline with realistic stages of use. Category guidance for skin care distinguishes delivery to the skin, pick up between the fingers, rub out on the application site, and after feel, with trained assessors rating the presence and evolution of attributes such as wetness, spreadability, thickness, oiliness, greasiness, and different residue types.

These standards also formalize panel leadership, reference anchors, and assessor selection and training so that measurements of intensity and duration are reproducible across sessions.Recent work in personal care extends Temporal Dominance of Sensations by pairing it with Temporal Dominance of Emotions in order to track how sensory trajectories coevolve with emotional responses during a single usage. Studies on facial creams demonstrate that temporal curves and combined sensory emotional trajectories differentiate products that otherwise appear similar when only mean liking is considered. Analytical workflows use principal component regression to link dominance rates for sensations to dominance rates for emotions across successive time intervals, thereby revealing how shifts in tactile or olfactory dominance correspond to changes in curiosity, satisfaction, or annoyance during pre application, application, and post application phases.The practical value of dynamic profiling for cosmetics is twofold. First, it captures critical moments when dominance switches from one attribute to another, for example from slipperiness during rub out to perceived residue during after feel, which often drive consumer judgments of comfort, cleanliness, and quality.

Second, it enables integration with hedonic endpoints so that development teams can diagnose whether a mismatch between early expectations and later sensations reduces liking, or conversely whether a progression toward positive sensations sustains engagement and improves preference. This dynamic perspective aligns with evidence that immediate sensory reactions shape compliance and that continued positive sensations support adherence to recommended use, both of which influence repeat purchase and brand loyalty for topical products. Although temporal methods originated in food research, their application to cosmetic and personal care products is now increasing. The literature describes that cosmetics involve multiple sensory modalities and that each modality responds to distinct formulation levers, which strengthens the rationale for temporal designs that follow the complete usage sequence rather than relying solely on static end point measurements. Dynamic profiling therefore complements threshold determinations by revealing when and for how long a stimulus reaches perceptual salience, and by connecting those temporal signatures to acceptance and rejection outcomes in real use.

III. CONCLUSION

Sensory thresholds give cosmetic development a precise language for what consumers can perceive and what they will accept. By treating detection and recognition as performance-based points on a psychometric function, and by expressing difference thresholds as just noticeable differences tied to a reference, we move beyond impressionistic vocabulary to numbers that guide real decisions. Acceptance and rejection thresholds complete that framework by locating the region where liking is protected and the point where it fails—turning perceptibility into market relevance.Method matters. Bias-resistant forced-choice designs provide stable estimates for detection and recognition. Difference testing converts small compositional or process shifts into user-relevant discriminability, allowing teams to set specification bands that prevent batch-to-batch drift from becoming noticeable. When these measurements sit alongside established descriptive programs for creams and lotions—staged across pick-up, rub-out, application, and after-feel—they translate cleanly into formulation targets, quality control limits, shelf-life guardrails, and defensible claims.

For practice, the path is straightforward: define the threshold construct that matches the decision at hand, choose a method that controls bias, embed testing within your category's descriptive framework, and express the outcome as actionable limits that engineering and manufacturing can hold. Maintain the link to consumers by pairing perceptual thresholds with acceptance diagnostics, so brand signatures remain identifiable yet comfortable across reformulation, scale-up, and storage. The current literature for cosmetics is still fragmented. Future work should standardize reporting of threshold protocols, extend temporal methods to capture how sensations and emotions evolve during real use, and leverage at-home or mobile testing to bridge laboratory control with everyday contexts. Stronger connections between thresholds, instrumental measures, and consumer segmentation will help tailor specifications to distinct markets without sacrificing reproducibility.Adopting this threshold-first approach yields products that are not only measurably perceivable and consistent, but also reliably liked. It equips development teams with a common quantitative framework, aligns R&D and manufacturing around clear guardrails, and ultimately supports faster iteration with fewer sensory surprises in market.

REFERENCES

[1] Ardoin, R., Romero, R., Marx, B., & Prinyawiwatkul, W. (2020). Exploring new and modified rejection-type thresholds using cricket snack crackers. *Foods*, *9*(10). https://doi.org/10.3390/foods9101352

[2] ASTM. (2015). Standard Guide for Two Sensory Descriptive Analysis Approaches for Skin Creams and Lotions 1. *Astm*, *i*.

[3] Bausenhart, K. M., Dyjas, O., Vorberg, D., & Ulrich, R. (2012). Estimating discrimination performance in two-alternative forced choice tasks: Routines for MATLAB and R. *Behavior Research Methods*, *44*(4). https://doi.org/10.3758/s13428-012-0207-z

[4] Bi, J., & Ennis, D. M. (1998). Sensory thresholds: Concepts and methods. ***Journal of Sensory Studies,*** *13*(2). https://doi.org/10.1111/j.1745-459X.1998.tb00079.x

[5] Deco, G., Scarano, L., & Soto-Faraco, S. (2007). Weber's law in decision making: Integrating behavioral data in

humans with a neurophysiological model. ***Journal of Neuroscience,*** *27*(42). https://doi.org/10.1523/JNEUROSCI.1072-07.2007

[6] Drake, M. A., Watson, M. E., & Liu, Y. (2023). Sensory Analysis and Consumer Preference: Best Practices. In *Annual Review of Food Science and Technology* (Vol. 14). https://doi.org/10.1146/annurev-food-060721-023619

[7] El Ouyoun Najm, N., Olabi, A., Kreyydieh, S., & Toufeili, I. (2010). Determination of visual detection thresholds of selected iron fortificants and formulation of iron-fortified pocket-type flat bread. ***Journal of Cereal Science,*** *51*(3). https://doi.org/10.1016/j.jcs.2009.12.004

[8] Filho, T. L., Della Lúcia, S. M., Minim, L. A., Cássia Silva, R. de, da Silva, A. N., & Rodrigues Minim, V. P. (2017). Validation of the Hedonic Threshold Methodology in Determining the Compromised Acceptance Threshold. ***Journal of Sensory Studies.*** https://doi.org/10.1111/joss.12255

[9] Filho, T. L., Della Lúcia, S. M., Minim, L. A., Cássia Silva, R. de, & Rodrigues Minim, V. P. (2018). Validation of the Hedonic Threshold Methodology in Determining the Hedonic Rejection Threshold. ***Journal of Sensory Studies***. https://doi.org/10.1111/joss.12313

[10] Gamba, M. M., Lima Filho, T., Torres, I. V., Della Lucia, S. M., & Minim, V. P. R. (2021). Random presentation minimizes the effect of expectation on the hedonic threshold methodology. *Food Quality and Preference*, *90*. https://doi.org/10.1016/j.foodqual.2020.104154

[11] García-Pérez, M. A., Alcalá-Quintana, R., Woods, R. L., & Peli, E. (2011). Psychometric functions for detection and discrimination with and without flankers. *Attention, Perception, and Psychophysics*, *73*(3). https://doi.org/10.3758/s13414-010-0080-8

[12] Guest, S., McGlone, F., Hopkinson, A., Schendel, Z. A., Blot, K., & Essick, G. (2013). Perceptual and Sensory-Functional Consequences of Skin Care Products. ***Journal of Cosmetics, Dermatological Sciences and Applications***, *03*(01). https://doi.org/10.4236/jcdsa.2013.31a010

[13] Hohl, K., Schonberger, G., & Busch-Stock, M. (2014). Stimulus and recognition thresholds for the basic tastes in deionized water: are the recommendations for citric acid too high? *Ernahrungs Umschau*, *61*(8).

[14] International Organization for Standardization. (2012). *Sensory analysis: General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors. ISO 8586:2012*.

[15] Keightley, C. J., Weare, A. M., Uriarte, M. De, Ma, X., Gent, U., Economie, F., Bedrijfskunde, E. N., Roman, K., Kroll, L., Annika Vignisdottir, H.M.U.S.R, S., K., B., N.S.D., C., R.P.K.N, R., P.A.K, D., Bruns, I., Liégeois, M., Rivera, C., Magnusson, A. T., … Lundin, P. (2016). Sensory marketing in natural cosmetics shops: the impact on generation X and generation Y. *International Business Research*, *11*(August).

[16] Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. In *Perception and Psychophysics* (Vol. 63, Issue 8). https://doi.org/10.3758/BF03194552

[17] Lawless, H. T., & Heymann, H. (2010). Sensory Evaluation of Food: Principles and Practices, 2nd Ed. In *Food Science Text Series*.

[18] Lima Filho, T., Minim, V. P. R., Silva, R. de C. dos S. N. da, Della Lucia, S. M., & Minim, L. A. (2015). Methodology for determination of two new sensory thresholds: Compromised acceptance threshold and rejection threshold. *Food Research International, 76*. https://doi.org/10.1016/j.foodres.2015.07.037

[19] Macmillan, N. A., & Creelman, C. D. (2004). Detection Theory: A User's Guide: 2nd edition. In *Detection Theory: A User's Guide: 2nd edition*. https://doi.org/10.4324/9781410611147

[20] Meilgaard, M. C. (1991). Testing for Sensory Threshold of Added Substances. ***Journal of the American Society of Brewing Chemists,*** *49*. https://doi.org/10.1094/asbj-49-0128

[21] Pensé-Lhéritier, A. (2015). Recent Developments in the Sensorial Assessment of Cosmetic Products: A Review. ***International Journal of Cosmetic Science***. https://doi.org/10.1111/ics.12223

[22] Theofanides, F., & Kerasidou, E. (2012). The Role of Sensory Stimuli on Perceived Quality of Cosmetic Products: An Empirical Examination of the Buying Process of Face Creams. *International Conference on …, Iccmi*.

[23] Trius-Soler, M., Santillán-Alarcón, D. A., Martínez-Huélamo, M., Lamuela-Raventós, R. M., & Moreno, J. J. (2020). Effect of physiological factors, pathologies, and acquired habits on the sweet taste threshold: A systematic review and meta-analysis. In *Comprehensive Reviews in Food Science and Food Safety* (Vol. 19, Issue 6). https://doi.org/10.1111/1541-4337.12643

[24] Webb, J., Bolhuis, D. P., Cicerale, S., Hayes, J. E., & Keast, R. (2015). The Relationships Between Common Measurements of Taste Function. *Chemosensory Perception*, *8*(1). https://doi.org/10.1007/s12078-015-9183-x.