IFSCC 2025 full paper (IFSCC2025-300)

"Development and Validation of a Noodling Atlas for Assessing Cosmetic Product Noodling"

Bing ZHANG[1], Lirong WEN[1], Chenqi CAO[1] and Gabriel CAZORLA [2,*]

[1] Beauty Innovation & Performance Center, Chanel China Co. Ltd., Shanghai, China; [2] Department of Research Biology, Chanel, Paris, France

1. Introduction

Cosmetic noodling is a commonly encountered issue in skincare and makeup routines, where small particles or strands form on the skin's surface during the product application and wearing process, resembling clumps or flakes [1]. This phenomenon not only hinders the proper absorption of the product, thereby compromising its effectiveness, but also negatively impacts user satisfaction, which can subsequently influence future product selection and usage. Recent studies have identified several contributing factors, including excessive product application, skin condition prior to use, improper layering of products, incompatibilities between the texture of the cosmetic and the skin's characteristics [1, 2].

Despite the widespread occurrence of noodling, the current assessment of noodling primarily relies on subjective perceptions and experiential judgments from consumers or professionals, which often lack consistency and reproducibility [3]. To address this issue, there is a pressing need for an objective, valid tool that can accurately quantify noodling severity. Such a tool would not only enhance the precision of product evaluations but also provide reliable data to support cosmetic research and development, offering valuable insights into product performance and its influence on consumer preferences. In response to this need, this study aims to develop a comprehensive Noodling Atlas that quantifies the degree of noodling in facial cosmetics. The atlas underwent rigorous validation by a team of internal project members with extensive industry experience, as well as dermatologists and professionals specialized in scale and atlas development. This collaborative process ensured the creation of a user-friendly and valid tool suitable for widespread application. Furthermore, to assess the practical utility of the Noodling Atlas in real-world cosmetic use, a questionnaire-based survey was conducted with a representative sample of Chinese cosmetic consumers, evaluating the acceptability of varying degrees of noodling severity.

2. Materials and Methods

2.1 Study design

This study employed two-phase design, consisting of the development and validation of the Noodling Atlas, and consumer acceptance testing. After obtaining the final atlas through

grade establishment and expert validation, the consumer testing was conducted on the atlas photos to understand how consumers accept the different grades of noodling in the atlas.

2.2 Development and validation of the Noodling Atlas

2.2.1 Development of the Noodling Atlas

2.2.1.1 Photo portfolio creation

A controlled photographic system was employed to ensure precise participant positioning and standardized lighting, critical for achieing consistency and reproducibility in image capture. The participant's face was illuminated using two Godox SK400II flash units, each set to half power, positioned to provide diffused light from fixed angles relative to the camera. This lighting configuration minimized shading and eliminated cast shadows, ensuring uniform ilumination across all images. Photography was conducted using a Nikon Z7II camera with a Z 105mm f/2.8 VR S lens, set to maual mode with an aperture of F/36, exposure time of 1/100 seonds, ISO 64, and a fixed distance of 50 cm between the camera and the participant's face. Color calibration was carried out before each session with a 48-color patch chart to maintain color accuracy.

The photographic range covers key areas of the face, including the cheeks, the lower edge of the eyes, part of the ears, and the chin. A trained technician captured high-resolution digital images, producing a total of 358 photos selected for inclusion in the atlas group. Images were taken with neutral facial expressions (i.e., no smiling, frowning, visible teeth, or partially open eyes), ensuring uniform magnification and consistent image size. The final photo portfolio and Noodling Atlas used in this study were based on images of a 33-year-old Chinese female participant with significant experience using cosmetic products, and no dermatologic or systemic conditions that could influence cosmetic performance. The collection comprised 357 photos representing varying degrees of noodling, alongside a baseline image (Grade 0) showing no visible noodling after product application.

2.2.1.2 Swiss system and grade establishment

Four internal project members, with extensive experience in the cosmetics industry, initially selected 100 photos from a pool of 357 images, aiming to identify those that most accurately represented varying degrees of noodling. Photos selected by at least two members advanced to the next phase, resulting in a subset of 128 images.

These 128 images were then subjected to the Swiss system, a structured evaluation process involving multiple rounds of pairwise comparisons [4]. After the seven rounds, the 128 photos were organized into 25 categories based on severity, with one or two images selected from each category to best represent the respective severity level, resulting in a final selection of 28 images. The 28 selected photos were then reviewed and discussed by the four project members. They were subsequently grouped into seven categories corresponding to different severity levels of noodling. Within each category, the most representative image was chosen, which was then paired with the baseline photograph (Grade 0) to form the initial version of the Noodling Atlas. A final re-evaluation of the images was conducted by the project members to confirm that all photos were accurately assigned to their respective severity grades (Grades 1-7).

2.2.2 Validation of the Noodling Atlas

Three external experts in the cosmetics field conducted two rounds of grading assessments on 227 photos selected from the photo portfolio. Expert 1, a dermatologist with 30 years of

professional experience, was joined by Expert 2 and Expert 3, both come from professional CROs (clinical research organization), with 4 and 9 years of experience in cosmetic efficacy assessment, respectively. Each expert performed two rounds of assessments with a minimum interval of 21 days between evaluations.

Due to the distribution of noodling severity within the Swiss system, which contained few examples of very low (Grade 1) and very high (Grade 7) noodling severity, all 7 images from these grades were included in the external validation set. For the remaining images, 110 were randomly selected from the 228 photos that had not previously been part of the Swiss system comparison, and another 110 were chosen from the 114 images within the Swiss system (excluding the 7 atlas photos and the 7 Grade 1 and Grade 7 images). To assess consistency, both intra-observer reliability (comparing first and second rounds of grading by the same expert) and inter-observer reliability (comparing results between experts) were evaluated.

2.3 Consumer testing

Consumers participanted in the consumer testing were recruited from various regions across China between March and April 2024.

2.3.1 Inclusion and exclusion criteria

The inclusion criteria for consumers in this study were as follows: (1) Chinese nationality, (2) female gender, (3) age range of 20 to 50 years, with an even distribution of participants (50%) in the 20-35-year age group and 50% in the 36-50-year age group, (4) use of premium skincare brands in the past 12 months, (5) residence in first- or second-tier cities in China, with a balanced representation from both tiers, ensuring coverage of eastern, southern, western, and northern regions of China. All participants provided informed consent prior to their inclusion in the study.

2.3.2 Consumer questionnaire

Consumer acceptance of various severities of noodling was evaluated through a questionnaire-based study. Participants were shown individual photos from the Noodling Atlas on screen, one at a time, and asked to assess whether they found the appearance of each image acceptable or unacceptable. To ensure the reliability and consistency of responses, each participant evaluated eight different photos in total.

2.4 Statistical analysis

During the validation process, photo presentation for scoring was randomized to mitigate potential bias. Reproducibility of the Noodling Atlas grading was assessed using two statistical methods. First, intra-observer agreement (consistency of the same observer across repeated assessments) and inter-observer repeatability (consistency across different observers) were quantified using the weighted Kappa coefficient [5]. Interpretation of the Kappa values followed the guidelines of Landis and Koch: values <0.2 indicated slight agreement, >0.2 to 0.4 fair agreement, >0.4 to 0.6 moderate agreement, >0.6 to 0.8 substantial agreement, and >0.8 to 1.0 almost perfect agreement [6]. Second, the degree of distinguishability (DD) between adjacent categories of the grading scale was assessed using a log-linear non-uniform association model [7]. DD values range from 0 (not distinguishable) to 1 (perfectly distinguishable), with values ≥0.6 indicating good distinguishability between grades. In cases where multiple observers had difficulty distinguishing between two adjacent grades, one of the grades was either excluded from the scale or its representative photo

replaced to ensure clarity. DD estimates were reported along with their 95% confidence intervals (CI), where applicable. All statistical analyses were performed using R software.

3. Results

3.1 Development of the Noodling Atlas

3.1.1 Swiss System

Following the seven rounds of the Swiss system, the 128 photos were categorized into 25 distinct groups based on their competition scores.

3.1.2 Noodling Atlas

The final Noodling Atlas developed in this study consists of eight photographs representing varying degrees of noodling severity, ranging from Grade 0 to Grade 7 (Figure 1). Grade 0 represents an image with no visible noodling, while Grade 7 depicts the most severe form of noodling observed in the study.

Figure 1. The Noodling Atlas

3.2 External validation of the Noodling Atlas

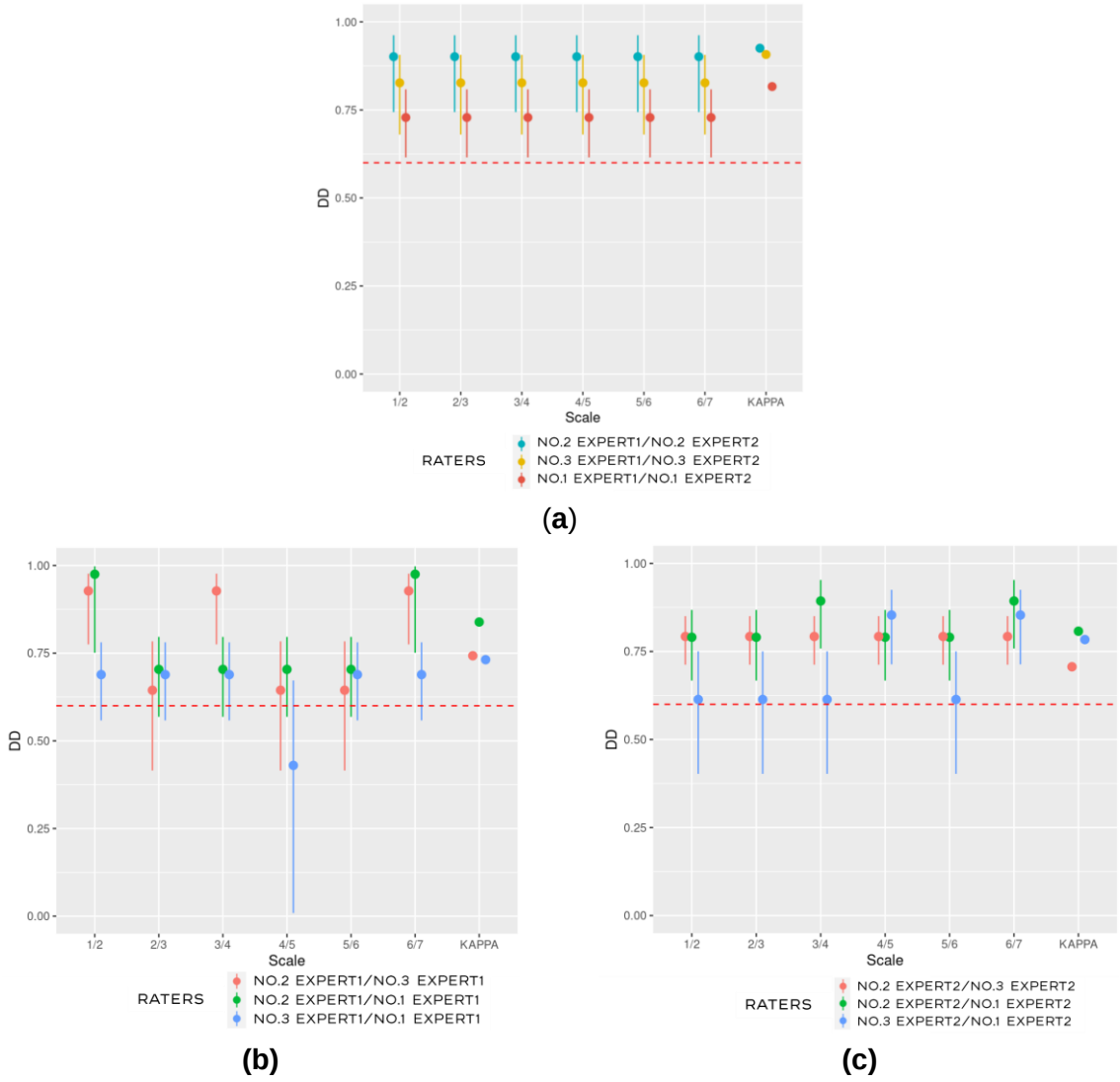
Inter-observer agreement and intra-observer reproducibility were assessed using the Noodling Atlas criteria. Agreement among the three experts across three rounds of evaluation on the 227 selected photos ranged from moderate to almost perfect, with weighted Kappa coefficients varying between 0.71 and 0.93 (Table 2).

Table 2. Intra-observer and inter-observer reliability

Weighted quadratic Kappa coefficient	Expert 1	Expert 2	Expert 3
Expert 1	0.82	0.84 (First round)	0.73 (First round)
Expert 2	0.81 (Second round)	0.93	0.74 (First round)
Expert 3	0.78 (Second round)	0.71 (Second round)	0.91

Kappa coefficient according to Landis & Koch [6]: <0.2 represents slight agreement; >0.2 to 0.4 fair agreement; >0.4 to 0.6 moderate agreement; >0.6 to 0.8 substantial agreement; and >0.8 to 0.9 strong agreement; >0.9 to 1.0 almost perfect agreement.

Intra-observer and inter-observer evaluations of the DD between adjacent grades showed very good agreement, with DD values exceeding 0.6 for most comparisons. The only exception was between Grade 4 and Grade 5 in the second round of evaluations by two raters (Figure 2). These results indicate that the grades in the Noodling Atlas are well-distinguished, and no revisions to the scale were necessary.



(a)

(b) **(c)**

Figure 2. Degree of distinguishability (DD) between adjacent categories in the external validation of Noodling Atlas. The DD values range from 0 (not distinguishable) to 1 (perfectly distinguishable), with DD values ⩾0.6 indicating good distinguishability between grades: (a) Intra-observer DD; (b) Inter-observer DD in the first round; (c) Inter-observer DD in the second round.

3.3 Consumer testing

3.3.1 Consumer characteristics

A total of 608 female cosmetic consumers from first- and second-tier cities in China participated in the questionnaire survey, with 598 valid responses collected. The average age of the participants was 32.25±7.12 years, with 340 (56.86%) in the 20-35 age group and 258 (43.13%) in the 36-50 age group. Geographically, the participants were distributed across various regions in China: 146 (24.46%) from the eastern region, 143 (23.95%) from the southern region, 155 (25.96%) from the western region, and 154 (25.63%) from the northern region. Of the total, 282 (47.16%) resided in first-tier cities, while 316 (52.84%) were from second-tier cities (Table 3).

Table 3. Consumer characteristics

Variables	Consumer (N=598)
Age (years), Mean±SD	32.25±7.12
Age group (years), n (%)	
20-35	340 (56.86)
36-50	258 (43.14)
Female, n (%)	598 (100)
Residential city, n (%)	
Tier 1 cities	282 (47.16)
Tier 2 cities	316 (52.84)
Residential regions, n (%)	
East	146 (24.41)
South	143 (23.91)
West	155 (25.92)
North	154 (25.75)

SD: standard deviation.

3.3.2 Consumer questionnaire

The questionnaire assessed consumer acceptability of noodling severity levels as represented in the Noodling Atlas. The results indicated high acceptability for lower severity grades, with 598 participants (100%) rating Grade 0 (no noodling) as acceptable. As the severity of noodling increased, the acceptability rates declined. Specifically, 587 participants (98.16%) found Grade 1 acceptable, while 579 (96.82%) accepted Grade 2. However, acceptability dropped significantly for higher severity grades, with only 471 (78.76%) accepting Grade 3, 368 (61.54%) accepting Grade 4, 188 (31.44%) accepting Grade 5, 49 (8.19%) accepting Grade 6, and only 26 (4.35%) accepting Grade 7 (Table 4).

Table 4. Consumer questionnaire acceptance

Photo grades, n (%)	Consumer (N=598)
Grade 0	598 (100.00)
Grade 1	587 (98.16)
Grade 2	579 (96.82)
Grade 3	471 (78.76)
Grade 4	368 (61.54)
Grade 5	188 (31.44)
Grade 6	49 (8.19)
Grade 7	26 (4.35)

4. Discussion

Our study successfully developed the Noodling Atlas, an objective tool for assessing varying degrees of facial cosmetic noodling. The four-stage selection process ensured a reliable representation of noodling severity, validated by external experts who confirmed strong consistency and clear distinguishability between grades. The consumer survey further highlighted the practical value of the Noodling Atlas, revealing a significant decline in acceptance as noodling severity increased. These findings underscore the utility of the Noodling Atlas as a valid and objective assessment tool, offering valuable insights for both cosmetic product development and consumer preferences.

The development of any evaluation tool for cosmetic efficacy or user experience must address consumer needs and accurately reflect their preferences. The Noodling Atlas developed in this study fulfills these criteria, as demonstrated by a representative survey of Chinese consumers. The results revealed a sharp decline in consumer acceptance as noodling severity increased, with 100% of participants accepting Grade 0 and only 4.35% accepting Grade 7, emphasizing the significant impact of noodling on user experience and its strong influence on purchase intentions. By providing an objective and standardized method for assessing noodling severity, the atlas offers considerable practical value. It establishes a

scientific basis for cosmetic development, enabling companies to address noodling effectively and improve product formulations.

Furthermore, the atlas has potential applications in product quality control, ensuring consistency across batches and combinations to maintain an acceptable level of noodling for consumers. These findings underscore the importance of integrating consumer-centric perspectives into the development of tools for cosmetic evaluation.

5. Conclusion

This study successfully developed and validated the first objective standard for evaluating the severity of noodling phenomena in facial cosmetics. The Noodling Atlas provides a reliable and valid tool for assessing noodling severity, addressing a critical gap in cosmetic evaluation. By establishing standardized grading criteria, this atlas offers valuable insights for cosmetic manufacturers and professionals, enabling them to improve product formulations and enhance user satisfaction.

Reference

1. Lua, B.L., et al., Understanding the causes of skincare product pilling. Skin Res Technol, 2024. 30(8): p. e13828.
2. Nayak, M., et al., Global Trends in Cosmetics Use-Related Adverse Effects: A Bibliometric Analysis of Literature Published during 1957–2021. Cosmetics, 2021. 8(3): p. 75.
3. Blaak, J., et al., Consumer panel size in sensory cosmetic product evaluation: A pilot study from a statistical point of view. Journal of Cosmetics, Dermatological Sciences and Applications, 2018. 8(3): p. 97-109.
4. Federation, F.I.C. General Rules and Technical Recommendations for Tournaments / 04. FIDE Swiss Rules / C.04.1 Basic rules for Swiss Systems. 2020; Available from: https://handbook.fide.com/chapter/C0401202507.
5. Cohen, J., Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull, 1968. 70(4): p. 213-20.
6. Landis, J.R. and G.G. Koch, The measurement of observer agreement for categorical data. Biometrics, 1977. 33(1): p. 159-74.
7. Valet, F., C. Guinot, and J.Y. Mary, Log-linear non-uniform association models for agreement between two ratings on an ordinal scale. Stat Med, 2007. 26(3): p. 647-62.