"Leveraging A/B testing in beauty-tech for advanced product development of personalized cosmetics"

Milène Chasseur[1], Anne-Sophie Brac De La Perriere[1], Camille Cléro[1], Thomas Navarro*[1], Romane Ricardou[1], Vincent Vandenbussche[1], Marie Mignon*[1]

[1]Science team Prose, Paris, France

1.   Introduction

In the beauty industry, formula renovations - whether aimed at reducing product cost, complying with evolving regulatory standards, or enhancing product efficacy - inevitably carry risks, particularly concerning how consumers will perceive the changes. In today's rapidly evolving beauty landscape, consumer expectations for personalized and effective products are higher than ever.

Traditionally, brands mitigate these risks through extensive pre-launch testing such as laboratory efficacy studies, expert panels, and consumer and clinical testing.

While these methods provide valuable insights, they often involve small sample sizes and controlled environments that may not fully reflect real-world consumer experiences.

To complement these traditional approaches, live A/B testing offers a novel, scalable solution to validate formula renovations under everyday usage conditions. Thus, live A/B testing offers a complementary and robust approach to de-risk formula renovations.

Widely adopted in tech and marketing for optimizing digital experiences, A/B testing involves exposing large groups of customers to different versions of a digital product (whether it is a webpage formats, button sizes, colors, copy, pictures, recommendations) without explicitly informing them that a test is underway, enabling data-driven continuous improvement [1].

Applied to physical beauty products, A/B testing entails testing different formula variations of a product directly with end consumers in their everyday lives, without disclosing that they are participating in an experimental comparison.

Despite formulation differences that may be observable on the INCI list, the test remains effectively blind, allowing brands to validate formula renovations with a higher degree of confidence than traditional methods alone.

This adaptation of A/B testing to physical product development represents a significant innovation in the beauty industry, offering a higher degree of confidence in decision-making than traditional methods alone.

This study aims to demonstrate how A/B testing can be effectively leveraged in the beauty industry to de-risk formula renovations by capturing real-world consumer feedback at scale. Two case studies are presented to illustrate the application of this methodology:

- Case Study 1 –Scalp Mask reformulation: Focused on reducing formula costs while maintaining consumer-perceived efficacy and overall satisfaction.
- Case Study 2 –Conditioner reformulation: Targeting improved performance on damaged hair through modification of the type of silicone used.

2.   Materials and Methods

The material and methods used in this study include the development of a dedicated A/B testing framework, comprehensive pre-test validation of renovated formulations, design and execution of live A/B tests, and post-test performance assessment through consumer feedback and business metrics.

2.1.   A/B Testing Framework Overview

We developed and validated an innovative A/B testing strategy designed to robustly confirm the superiority or non-inferiority of renovated formulas in terms of consumer satisfaction and perceived performance under real-world conditions. This approach complements traditional pre-launch evaluation methods, including expert panel testing and at-home use testing, which provide initial insights into product performance, safety, and consumer acceptance. Our A/B testing framework adds a critical layer of in-market validation by assessing consumer satisfaction at scale and under natural usage conditions.

2.2.   Pre-A/B test formulation validation

Prior to live testing, the renovated versions of both the scalp mask and conditioner underwent comprehensive internal validation, including assessments of efficacy, safety, quality, and regulatory compliance. The products moved into A/B testing phase only after successfully meeting all validation criteria.

Specifically, the scalp mask reformulation significantly reduced formula cost while maintaining equivalent performance based on expert assessments and at-home use testing. For the scalp mask, the cost optimization strategy focused on reformulating the product with more cost-effective raw materials and process efficiencies, while maintaining the overall sensorial properties, efficacy, and compliance with quality standards. No significant changes in active ingredients or key functional components were introduced to preserve the original product performance profile.

The conditioner reformulation, incorporating a different type of silicone, delivered enhanced conditioning benefits, particularly on bleached hair, as confirmed by expert and consumer evaluations.

2.3.   A/B Test Design

During the purchase of a scalp mask or conditioner on the brand's e-commerce platform consumers were randomly assigned to receive either Formula A (the original product, control) or Formula B (the renovated variant, test). The test population included both new and returning customers to ensure a representative sample. Consumers were blinded to their group assignment, thereby ensuring unbiased consumer perception and natural usage behavior.

The scalp mask A/B test was designed with a 50%/50% allocation ratio between the control and test formulas, while the conditioner A/B test was designed with a 70%/30% allocation ratio in favor of the control formula. Actual sample sizes and order volumes are presented in the Results section.

2.4. Performance Assessment
2.4.1. Satisfaction Metrics

A structured post-use feedback loop was embedded into the customer journey. Three weeks after product shipment, consumers received an email prompting them to complete a satisfaction survey hosted on the brand's website, aimed at reviewing product performance and enabling potential formula refinement for future orders. The survey collected overall satisfaction and key performance attributes (e.g., efficacy, sensorial experience, conditioning effect) rated on a 1–5 scale.

Collected data were analyzed using two complementary statistical approaches: Chi-square tests comparing Top2Box (ratings of 4 or 5) and Bottom2Box (ratings of 1 or 2), with a significance threshold set at $p < 0.05$ and a Bayesian statistical analysis to strengthen the robustness of conclusions. Error bars in figures represent 95% confidence intervals.

2.4.2. Business Metrics

Two key business KPIs were tracked and compared across control and test groups to measure the business impact of the formula changes :

- **Reordering rate**: defined as the percentage of consumers who placed a repeat order of the product within 6 months following their initial order placed during the A/B test.
- **Life Time Value (LTV)**: defined as the total revenue generated by a customer starting from the A/B test order over the 12 following months. LTV was analyzed only in cases where no significant difference in reorder rate was observed, to assess long-term value beyond immediate repurchase.

3. Results

A total of 96,222 orders were included in the scalp mask A/B test, with an intended 50%/50% allocation between control and test formulas. For the conditioner A/B test, 652,711 orders were included, with a 70%/30% control-to-test allocation ratio. Consumer feedback was collected via satisfaction surveys, resulting in 1,760 responses for the scalp mask (approximately 1.8% response rate) and 82,447 responses for the conditioner (approximately 12.6% response rate). The analyses below present the comparative results of the control and renovated formulas based on satisfaction metrics and business performance indicators.

3.1. Scalp Mask A/B Test
3.1.1. Satisfaction Metrics

A non-superiority chi-square test showed that the proportion of dissatisfied users (Bottom2Box) for the renovated formula was not significantly higher than that of the control formula, confirming that the cost-optimized version did not increase dissatisfaction (Figure 1).

A non-inferiority chi-square test demonstrated that the proportion of satisfied users (Top2Box) for the renovated formula was not significantly lower than for the control. This supports the conclusion that the renovated formula maintained comparable satisfaction levels (Figure 1).
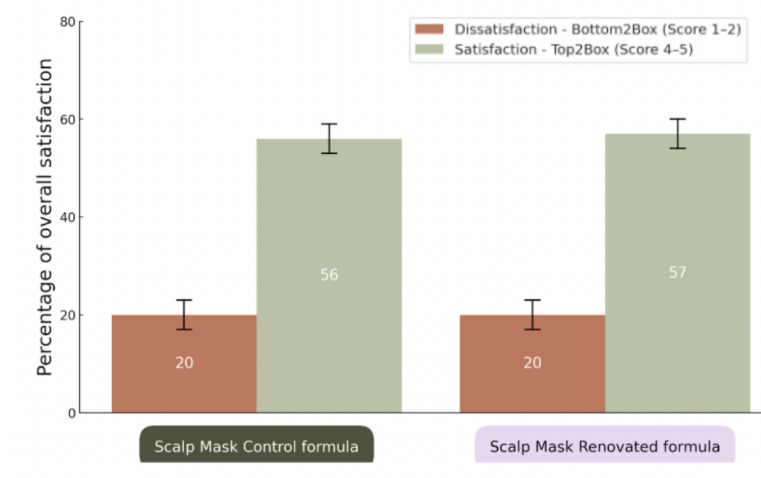

Figure 1. Comparison of product satisfaction (Top2Box and Bottom2Box scores) between control and renovated scalp mask formulas.

The cost-optimized scalp mask formulas were statistically shown to be non-inferior to the Control in terms of product satisfaction, and non-superior in terms of product dissatisfaction. This validates the reformulation as successful, confirming that cost savings were achieved without negatively impacting the consumer experience. Furthermore, satisfaction with specific benefits—such as improved scalp health, reduced flakiness, decreased oiliness, and overall scalp comfort—was also statistically non-inferior for the renovated Scalp Mask formula compared to the control formula. This consistency across individual benefits further reinforces the overall success of the renovation.

3.1.2. Business Metrics

The percentage of consumers who reordered the scalp mask following their A/B test order was similar between the control and test groups, indicating no negative impact on repurchase behavior (Figure 2).
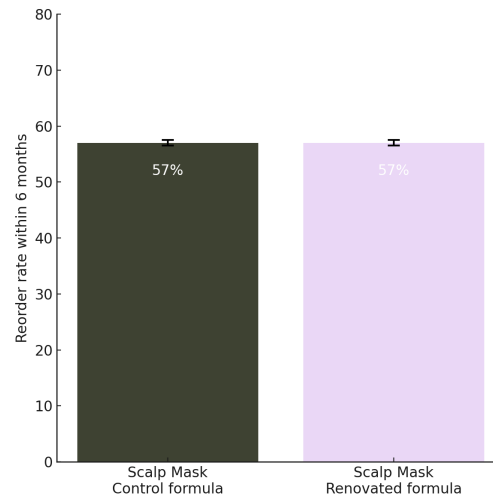
Figure 2. Comparison of reorder rates between control and renovated scalp mask formulas within 6 months following the initial A/B test order

3.1.3. Overall Impact

Both satisfaction and business metrics confirmed that the introduction of the renovated, cost-optimized scalp mask formula had no negative effect on the consumer experience. This enables potential margin optimization without compromising consumer satisfaction, perceived performance, or loyalty.

3.2. Conditioner A/B Test
3.2.1. Satisfaction Metrics
3.2.1.1. Product Satisfaction

A chi-square test of difference confirmed that the renovated formula demonstrated statistically significant improvements, with a higher proportion of satisfied consumers (Top2Box) and a lower proportion of dissatisfied consumers (Bottom2Box) compared to the control (Figure 3).
A Bayesian posterior probability analysis further supported these results, showing a 99.7% probability that the test group had a lower proportion of dissatisfied consumers (Bottom2Box) than the control, and a 100% probability that the test group had a higher proportion of satisfied consumers (Top2Box) compared to the control group.
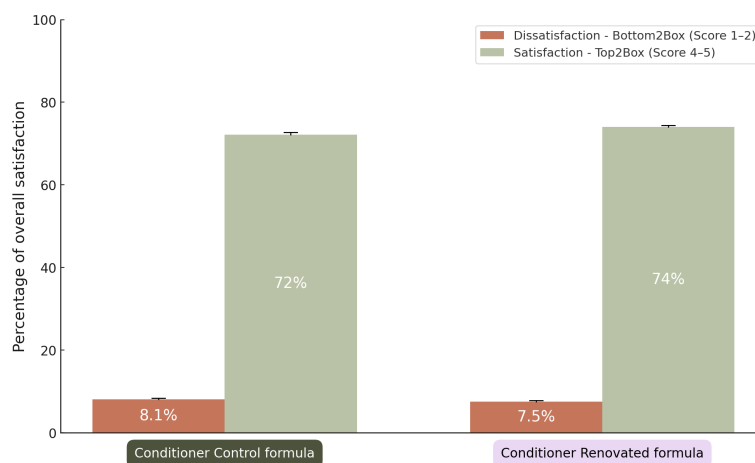
Figure 3. Comparison of overall satisfaction (Top2Box and Bottom2Box scores) between control and renovated conditioner formulas.

These findings provide strong evidence that the renovated conditioner significantly improved overall consumer satisfaction.

3.2.1.2. Product Benefits

The use of a new silicone type in the conditioner formulation resulted in a significant reduction in the number of consumers reporting an under-conditioned feel, and a significantly increased number of consumers reporting soft and manageable hair. These improvements reinforce the superior conditioning performance of the renovated formulation (Figure 4).
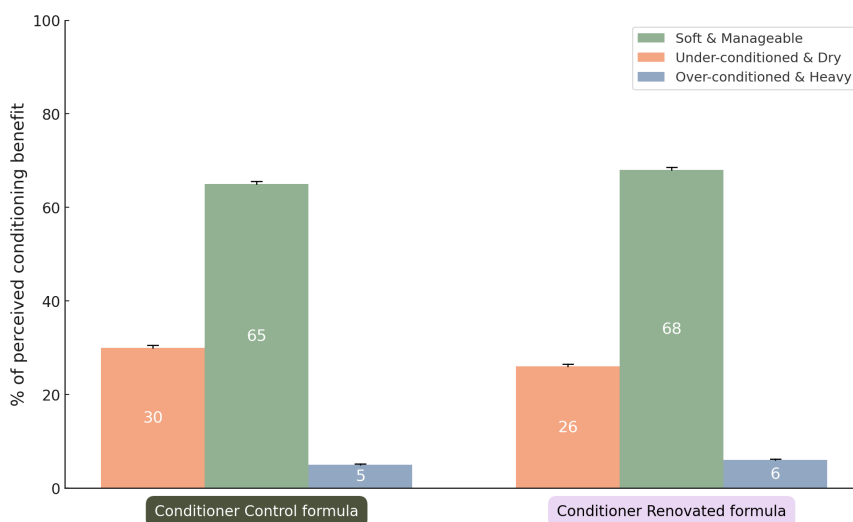

Figure 4. Comparison of perceived conditioning benefits between control and renovated conditioner formulations.

3.2.2. Business Metrics

The A/B test did not result in a statistically significant difference in reorder rates between the control and test groups for the conditioner. However, a statistically significant uplift in 12-month Life Time Value (LTV) was observed for consumers who received the renovated formula, across both first-time buyers and returning customers (Table 1). This indicates that while repurchase frequency remained stable, the overall customer value increased with the new formula.

	12-Month Life Time Value (LTV)	
	Conditioner Control Formula	Conditioner Renovated Formula
New Customers	298.9$	301.2$
Returning Customers	291.3$	302.5$

Table 1. Comparison of 12-Month Life Time Value (LTV) between control and test groups for the conditioner.

3.2.3. Overall Impact

Both satisfaction and business performance indicators confirm that the renovated conditioner formula positively impacted the consumer experience and overall business value,validating the success of the renovation.

4.   Discussion

This study demonstrates that A/B testing is an innovative consumer-centric methodology that can complement traditional efficacy testing in the beauty industry. By capturing real-world feedback at scale under natural usage conditions, A/B testing provides actionable insights that validate product changes with greater confidence and consumer relevance.

Our results showed that the cost-optimized scalp mask formula maintained equivalent consumer satisfaction and business performance compared to the original version, enabling potential margin improvements without compromising user experience.

Similarly, the renovated conditioner, incorporating a new silicone type, significantly improved consumer satisfaction and positively impacted customer lifetime value, demonstrating the capacity of A/B testing to detect meaningful enhancements in product perception and business metrics. Beyond evaluating formulation or ingredient modifications, the versatility of A/B testing could extend to a wide range of product renovations scenarios, including texture refinements, manufacturing process optimizations, packaging improvements, and more. This adaptability positions A/B testing as a powerful tool for driving continuous innovation and elevating

consumer satisfaction, particularly in the beauty e-commerce space where consumer expectations evolve rapidly.

However, this study also has limitations. Despite blinding, consumers theoretically had access to formulation details (via INCI lists) which could, albeit unlikely, influence perceptions.

Finally, this methodology captures short- to mid-term satisfaction; longitudinal studies would be necessary to assess sustained consumer loyalty and product performance over time.

In conclusion, A/B testing offers a robust, scalable, and consumer-focused methodology that complements traditional product development processes. It enables beauty brands to innovate confidently while maintaining or enhancing consumer trust and satisfaction.

5. Conclusion

These studies demonstrate the effective application of A/B testing in cosmetic product development, illustrated through two distinct use cases:

- The refinement of a scalp mask formula aimed at reducing costs while preserving efficacy and maintaining high consumer satisfaction,

- The renovation of a conditioner involving a silicone substitution to enhance conditioning performance, particularly on damaged hair.

Through real-world consumer feedback collected at scale, A/B testing enabled robust validation of product changes, providing a powerful complement to traditional laboratory and panel testing methods.

This approach allowed for confident, data-driven decision-making while minimizing risks associated with formula renovations.

More broadly, these findings highlight the potential of A/B testing as a strategic tool for continuous product improvement and innovation in the beauty industry, particularly within e-commerce contexts where rapid iteration and consumer-centric development are increasingly critical.

References

[1] Quin, F., Weyns, D., Galster, M., & Costa Silva, C. (2024). A/B testing: A systematic literature review. *The Journal of Systems and Software, 211*, 112011.