

**UNCOVERING HIDDEN LINKS:
USING SENSORY DATA TO
PREDICT CONSUMER
BEHAVIOR WITH MASCARA**

Natalie Armstrong

MARY KAY

Quirks – Dallas
February 28, 2024



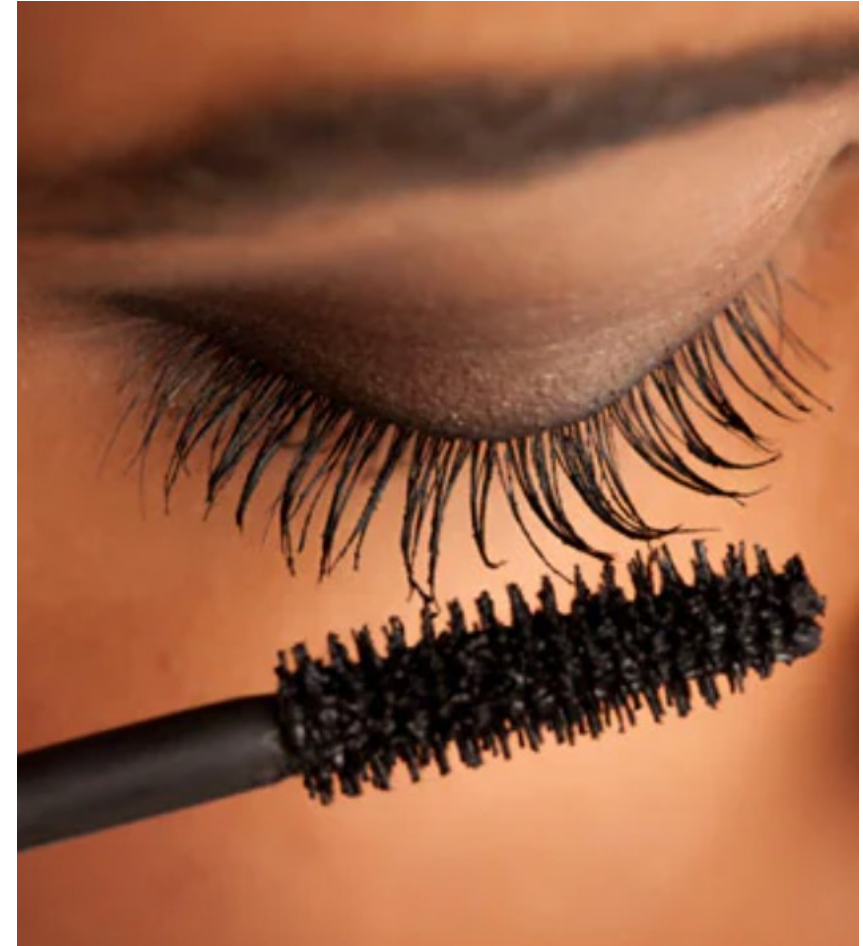
OVERVIEW

- Background, Objectives, and Analytic Plan
- **Phase 1:** Consumer home-use test (HUT) results
- **Phase 2:** Sensory testing & perceptual mapping results
- **Phase 3:** *NEW* Consumer home-use test (HUT) results
- Conclusion & Key Takeaways

HOW DID WE GET HERE

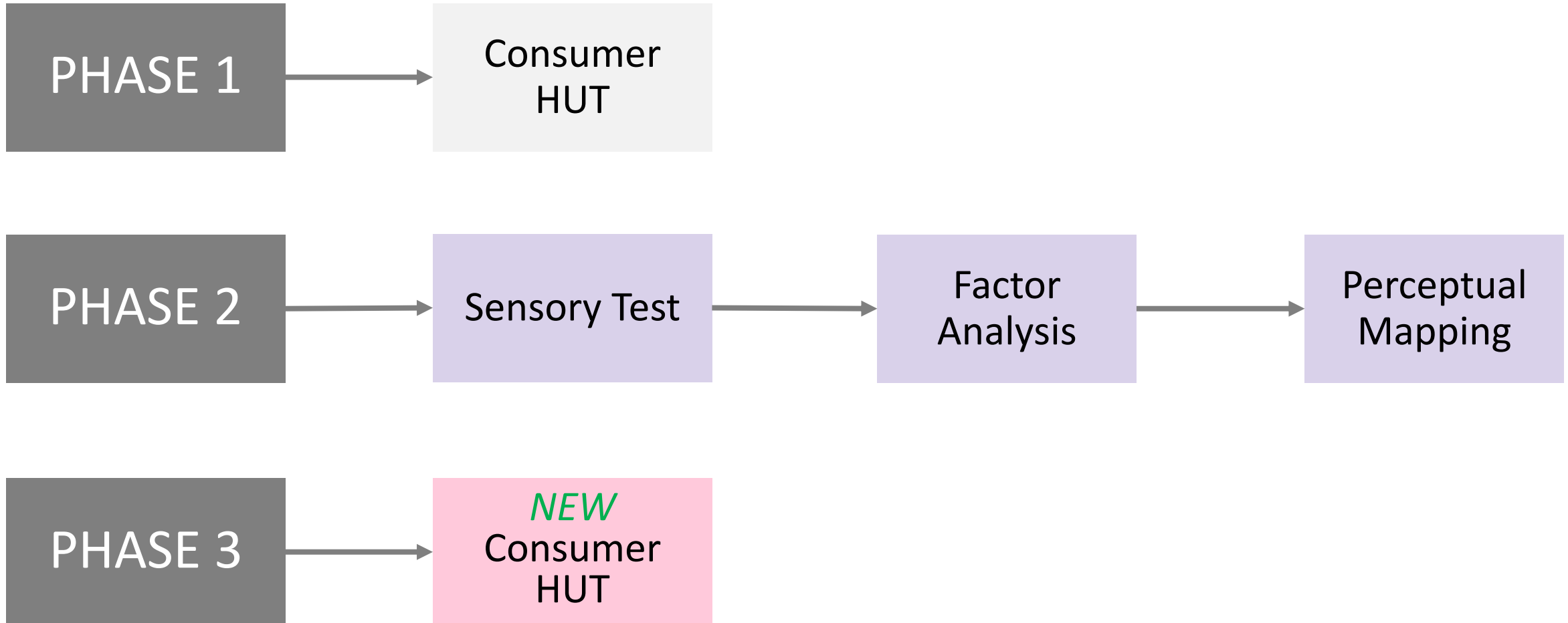
BACKGROUND

- Reviewing the consumer results from both internal prototypes and competitors indicated **little consumer differences between products.**
- The lack of differentiation between test products revealed the need to conduct sensory testing and combine this data with consumer results for a **comprehensive mixed methodology approach.**



OBJECTIVES & ANALYTICAL PLAN

- **Phase 1** – Understand consumer preference and liking among prototypes and competitors.
- **Phase 2** – Analyze differences among test samples using sensory test and perceptual mapping.
- **Phase 3** – Identify a top prototype based on dimensions discovered in phase 2.



PHASE 1: CONSUMER HOME-USE TEST (HUT) RESULTS

STUDY DESIGN & FIELDWORK (CONSUMER HUT)

METHOD

- Prototypes were unbranded (n=4), and competitors were masked (n=4).
- To prevent bias, all test samples were rotated within the study design – **incomplete block design**.
- 4-day usage study (Home-Use Test)
- **End-consumers** completed a survey on agreeance statements after using the mascara.
 - Statements focused on **application, appearance, mascara performance, wear/durability, etc.**

SAMPLE

- End consumers who purchased mascara recently.
- Frequent mascara users
- Mix of different types of lashes (limit long & thick lash mascara users)
- Sample size: N = 100
- **Total Number of Completes per Product: 50+**

CONSUMER HUT RESULTS

- Looking at the few statistically significant differences and the low magnitude of the scores, no prototype was ideal to move forward with.
- Of particular interest was **Overall Liking** and **Appearance** agreement statements but given the flat data, it was difficult to provide insights that could be activated.

	Prototypes				Competitors			
Agreement statements* Scores reflect % agreement with each statement	Prototype 1 (A)	Prototype 2 (B)	Prototype 3 (C)	Prototype 4 (D)	Competitor 1 (E)	Competitor 2 (F)	Competitor 3 (G)	Competitor 4 (H)
Overall Liking (Range: 0 low – 9 high)	6.0	6.7	7.0 A	6.7	6.9 A	6.6	7.1 A	6.4
Application statement	45%	52%	58%	52%	50%	46%	63% af	49%
Brush statement	79%	87%	81%	85%	85%	77%	77%	79%
Appearance statement	76%	79%	83%	81%	79%	77%	81%	71%
Performance statement	60%	65%	71% a	69%	68%	69%	72% a	65%
Removal statement	75%	85%	81%	71%	73%	77%	77%	77%
Wear/Durability statement	48%	48%	60% f	58% f	44%	42%	58%	48%

*Alpha Coding at the 95% level (UPPER CASE) and 90% level (lower case)

CONSUMER STATEMENTS CORRELATION WITH OVERALL LIKING

- Analysis did consider all statements, however, those listed in the table below are the top statement's Correlation scores with **Overall Liking**.
- Correlation results indicate that **Appearance** and **Performance** statements were the primary drivers of **Overall Liking** scores.
- Thus, despite relatively flat consumer data, insights were provided to focus on these specific **Appearance** and **Performance** for future prototype iterations.

Agreement Statements	Correlation Score w/ Overall Liking
Appearance statement	0.72
Performance statement	0.68
Brush statement	0.60
Application Statement	0.60
Appearance statement 2	0.57
Application Statement 2	0.54
Appearance Statement 3	0.52
Wear/Durability Statement	0.49

Meaningful cut in correlation scores placed here at 0.49 since all other statement's correlation scores were 0.30 and lower (removed from this table).

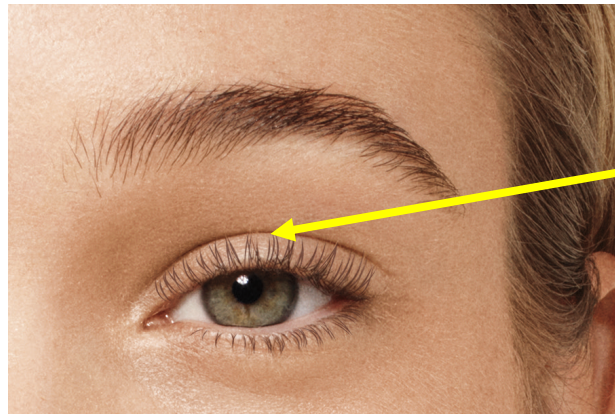
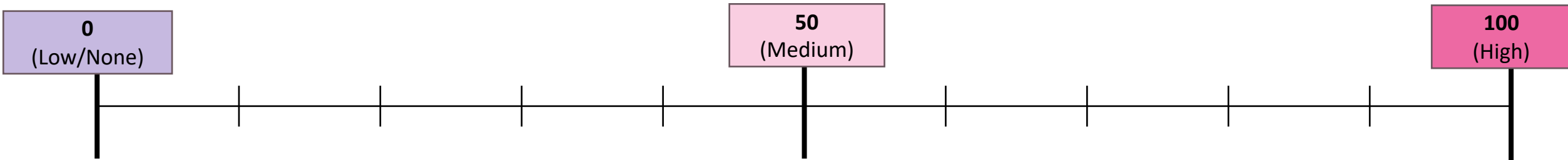
PHASE 2: SENSORY TESTING & PERCEPTUAL MAPPING

STUDY DESIGN & FIELDWORK (SENSORY – EXPERT TRAINED PANEL)

- Sensory testing is ideal to understand various small differences among test products.
- 9-12 expert panelist, trained on Descriptive analysis completed the study.
- Samples were randomized among panelist, with each product evaluated twice.
 - The same products tested for the consumer HUT study were included in the sensory testing study.
- Ballot had 30 product attributes.
 - A 0-100 intensity scale was used to evaluate the attributes
 - Scores range in intensity 0 (low/none) - 100 (high).
 - Types of product attributes:
 - Application, performance, appearance attributes (e.g., length, smudging, ease of application, etc.)
- Application was uniform across panelist.
- Data was summarized for each attribute among each product.
- Statistical procedures (Tukey's ANOVA) are used to highlight significant differences among samples for each.

EXAMPLE OF EXPERT PANEL TRAINING

- Pictures are provided to panelists during training to show the differences between low, medium, and high intensity for each attribute.
- Panelists must then calibrate with each other among various attributes to maintain consistency across the panel.



EXAMPLE: Short lashes

EXAMPLE: Long lashes



SENSORY DATA RESULTS

- As expected, there were small statistical differences found among test products, but multiple small differences can have an additive effect that can lead to an overall difference between products.
- The upcoming perceptual mapping analysis proved essential here to help visualize the cumulative effect of these small differences.

Attribute* <i>Scores range in intensity 0 (low/none) - 100 (high)</i>	Prototype 1	Prototype 2	Prototype 3	Prototype 4	Competitor 1	Competitor 2	Competitor 3	Competitor 4	Sig
Application Attribute 1	57.7 C	59.1 B	61.3 AB	58.1 B	54.9 D	59.7 B	57.9 B	60.3 A	**
Application Attribute 2	61.5	61.6	61.3	61.3	61.1	60.6	61.7	61.6	NS
Appearance Attribute 1	60.4 B	60.8 B	58.9 B	58.4 B	58.4 B	60.3 B	63.8 A	60.2 B	*
Appearance Attribute 2	67.4 B	68.6 AB	68.4 AB	66.6 B	68.5 AB	66.7 B	70.2 A	67.6 B	**
Performance Attribute 1	73.8 A	74.1 A	74.4 A	73.6 A	73.9 A	73.0 AB	71.4 B	73.2 AB	*

*Means that share a common letter within an attribute ARE statistically SIMILAR; * = .10, ** = .05; NS = Not statistically significant

SENSORY PERCEPTUAL MAPPING ANALYSIS

PERCEPTUAL MAPPING

- Summarize similarities and differences among products
- Visually represent the sensory space of a category

METHODOLOGY – FACTOR ANALYSIS

- **Why perform factor analysis?**
- Factor analysis is a powerful method that **shrinks/reduces the mass data** into smaller datasets/categories (i.e., dimensions) making the data more manageable and understandable.
 - It is a way to **find hidden and intricate patterns**
 - Shows how **dimensions/patterns overlap**
- Key dimensions were then taken and plotted on a perceptual map.

RESULTS – FACTOR ANALYSIS

- **Dimension 1** dominates data variability at 43.4 % and loads attributes that represent lash appearance and lash application, **Dimension 2** accounts for 22.1% of data variability and includes lash performance and removal attributes.
- Despite other dimensions, these dimensions are the top dimensions that are driving the differences and similarities between the test products.

Dimension 1

(43.4% of variability)

Appearance attributes
Application attributes

Dimension 2

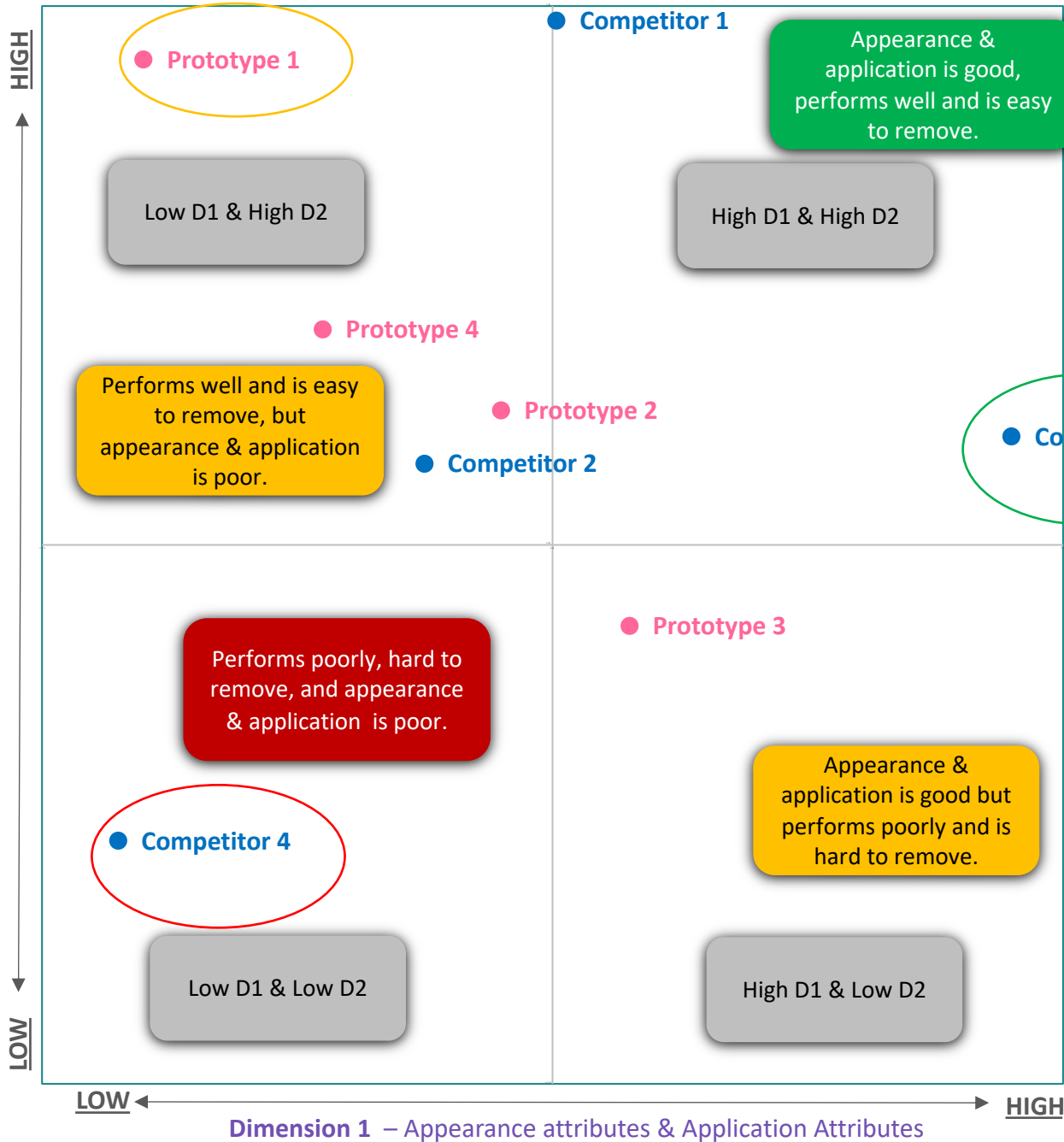
(22.1 % of variability)

Removal attributes
Performance attributes

Dim 1 vs. Dim 2

(Combined 65.5% of the variability)

Dimension 2 : Performance attributes & Removal attributes



Low D1 & High D2: **Prototype 1** is highly defined in this quadrant and while this prototype performs well throughout the day, appearance and application is poor.

High D1 & High D2 [IDEAL]: Products in this quadrant are high in both Dimension 1 & Dimension 2 attributes. Thus, these products look and apply the best and also perform well throughout the day. **Competitor 3** is the only product in this category.

Low D1 & Low D2 [WORST]: Products in this quadrant are low in both Dimension 1 and Dimension 2. The mascaras perform poorly and are also plagued with application issues and poor lash appearance. **Competitor 4** is the only product in this category but is not highly defined by this quadrant.

High D1 & Low D2: While application and appearance is good with products in this category, performance and removal will be poor. No mascara is highly defined in this quadrant.

KEY TAKEAWAY: Most prototypes were located in the middle of the graph indicating overall that we were in the right direction, but improvements could be made to appearance and application.

NOTE: Mock data is presented as an example and does not represent actual scores from testing.

**PHASE 3: *NEW* CONSUMER HOME-USE
TEST (HUT) RESULTS**

NEXT STEPS FOR *NEW* CONSUMER TESTING

- **QUICK RECAP (Phase 1 & 2)**

- **Consumer HUT results** – Top drivers of Overall Liking were Appearance and Performance statement
- **Sensory results** – Where our prototypes were located when compared to competitors and what attributes to focus on to improve formulas.

NEXT STEPS:

- Prototypes were reformulated based on insights provided.
- Consumer testing for new prototypes was overhauled:
 - New objective
 - New success criteria
 - Aligned sensory attributes (phase 2) w/ consumer statements
 - Revised the study design to be a preference study

STUDY DESIGN & FIELDWORK

(*NEW* CONSUMER HUT)

METHOD

- 4x prototypes and a control were unbranded and masked (no competitors included).
- Prototypes were directly compared and paired with the control.
- To prevent bias, all test samples were rotated within the study design – **incomplete block design**.
- 4-day usage study (Home-Use Test)
- **End-consumers** completed a survey on sample preference to specific statements after using the mascara.
 - Statements focused on **application, appearance, mascara performance, wear/durability, etc.**

SAMPLE

- End consumers who purchased mascara recently.
- Frequent mascara users
- Mix of different types of lashes (limit long & thick lash mascara users)
- Sample size: N = 100
- **Total Number of Completes per Product: 50+**

NEW CONSUMER HUT PREFERENCE RESULTS

- While Prototype 1 stands out as it was statistically preferred by consumers for most statements, other prototypes mostly performed similar to the control across preference statements.
- Also, consumer overall preference indicates that Prototype 1 was the only test prototype to achieve statistically higher preference than the control.

Recommended prototype

Preference Statements* *Scores indicate percent of panelists that prefer the test product	Prototype 1 (A)	Control (B)	Prototype 2 (A)	Control (B)	Prototype 3 (A)	Control (B)	Prototype 4 (A)	Control (B)
Overall Preference	62% b	38%	54%	46%	61%	39%	55%	45%
Application Statement 1	62% b	38%	56%	44%	58%	42%	48%	52%
Application Statement 2	60%	40%	67% B	33%	63% b	37%	53%	47%
Appearance Statement 1	67% B	33%	65% B	35%	56%	44%	53%	47%
Appearance Statement 2	64% B	36%	53%	47%	60%	40%	53%	47%
Appearance Statement 3	66% B	34%	67% B	33%	58%	42%	58%	42%
Performance Statement 1	69% B	31%	57%	43%	60%	40%	57%	43%

*Alpha Coding at the 95% level (UPPER CASE) and 90% level (lower case)

CONCLUSION

KEY TAKEAWAYS

This research was able to uncover hidden links between mascara product attributes among otherwise flat consumer data.

By combining methodologies to connect consumer data to sensory product attributes, we were able to unlock powerful insights to provide guidance on future mascara iterations.

This method was ultimately successful as future consumer testing indicated positive consumer reception.

