CONCLAVO

ORIGINAL RESEARCH



CONCLAVO RESEARCH BRIEF

What should Amazon sellers test for AI shoppers?

By Conclavo

X @Conclavo_

Conclavo Strategy Room

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

THE SELLER'S DECISION

What could rule your product out?

We asked an AI shopper to find a shower cap that would keep a buyer's long hair dry. A low price helped only if the cap looked big enough to do the job.

BUYER NEEDS

Room for long hair

Will the cap cover it and still seal around the hairline?

SELLER OFFERS

A $4.08 cap

Does the listing show why this particular cap will fit?

The seller's decision was whether to change the product, lower the price or make its capacity clearer.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

FIRST RECOMMENDATION

Test the fact your buyer needs first.

Pick one buying situation and one rival. Change a single truthful detail that could determine whether your offer qualifies.

Fit is uncertain

Show a verified size, capacity or compatible model where the buyer first compares offers.

Quantity is fixed

Compare the full order cost at the quantity the buyer needs, including mixed packs.

Performance is disputed

Check the recurring complaint against the product. Fix a real defect or state a precise limit.

CONCLAVO'S GUIDANCE

Change one thing. Count how often the revised offer wins against the same rivals, then repeat with fresh shoppers.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

STUDY DESIGN

How we did this

Fresh AI shoppers using Codex chose products from private copies of real Amazon listings. We changed what one shopper could see or buy and recorded its recommendation.

1,740

Tape and batteries

Proposed baskets, pack sizes, prices and ratings.

1,640

Reusable shower caps

Fit, search, reviews, competitors and repeat decisions.

3,380 shopping decisions across three product categories. A second look at an existing decision was counted only once. No purchases were made.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

EXECUTIVE SUMMARY

Which changes won more recommendations?

Add a missing size fact

1 to 48

Blue Dot wins out of 48 per version, with four caps visible.

Show a higher star average

56 to 64

Blue Dot wins out of 64 per version, with a privately edited 4.1 or 4.7 stars.

Move existing fit wording

43 to 45

Wins out of 192 per version across three sellers.

Describe a photo link

47 to 47

Blue Dot wins out of 48 per version.

CONCLAVO'S GUIDANCE

Start with missing information that could change the buyer's decision. An extra detail about an already credible offer may do little.

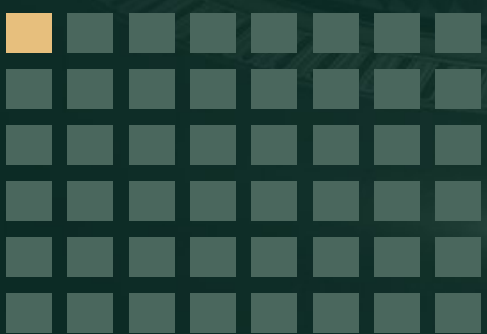
© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

TITLE TEST · FOUR COMPETING CAPS

“Extra large” changed the winner.

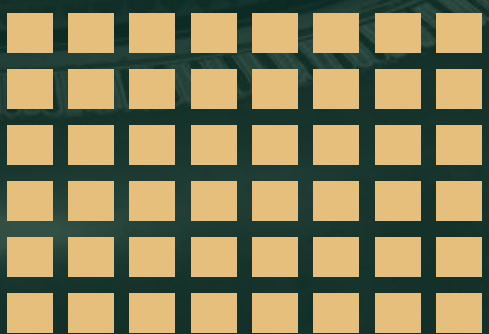
The Blue Dot cap stayed at $4.08. Only capacity and wording about staying on the head varied in the title. Each square below is one recommendation.

WITHOUT THE SIZE WORDING



1 of 48

WITH THE SIZE WORDING



48 of 48

A second phrase, “won’t fall off your head,” moved Blue Dot from 24/48 to 25/48. Both changes sounded relevant. Only the size wording made a large difference here.

CONCLAVO’S GUIDANCE

Use the title to answer the buyer’s fit question with a fact you can substantiate.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

PACK-SIZE TEST · BROTHER LABEL TAPE

Price the whole order.

A buyer needed at least seven cartridges. A single cartridge cost more per unit than a double pack, yet adding the single produced the cheaper order.



$91.45

Three doubles + one single



$99.96

Four doubles

The mixed order saved $8.51. All 8 shoppers chose it. When the buyer needed six or eight cartridges, all 16 chose doubles alone.

CONCLAVO'S GUIDANCE

Calculate the cheapest basket at your customers' common quantities before changing the price per unit.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

BEST-VALUE BATTERY TEST

Lower stars drew comments. The cheapest basket kept winning.

Buyers needed at least 40 batteries. The cheapest basket included a 24-pack. Lowering its rating gave Codex a reason to discuss a rival.

MENTIONED RATINGS

1 / 192 → **120 / 192**

CHOSE CHEAPEST BASKET

192 / 192 → **191 / 192**

Left: 4.8 stars. Right: 4.3 stars. The other packs stayed at 4.8. The cheapest order cost $37.52 in both versions.

CONCLAVO'S GUIDANCE

Track the basket an agent recommends separately from the concerns it mentions afterward.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

SHOPPING BEHAVIOR · 24 REAL LISTINGS

Codex saw every listing. It opened only some.

An appearance in search gave each seller a chance. Asking for a comparison changed how many products got a closer look.

ORDINARY REQUEST



9 products inspected, including the winner

ASKED TO COMPARE



11 products inspected, including the winner

Each row shows the middle shopper’s inspection count. All 96 shoppers saw all 24 cards. Comparison shoppers received more review texts on average, while fewer opened a reviewer photo (18/48 versus 38/48).

CONCLAVO’S GUIDANCE

Track whether your product appeared, whether Codex inspected it, and whether it recommended it. Each answers a different seller question.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

REVIEW EVIDENCE · 78 PAIRS WITH DIFFERENT WINNERS

Both shoppers pointed to the same drying problem.

Both shoppers often found the same reason to pass over one cap, then recommended different alternatives.

63 of 78

Both shoppers cited ELEBOX's drying, odor or moisture problem and received at least one identical review supporting it.

After the study we checked this recurring complaint in pairs that picked different winners. We did not test a fix to the cap.

CONCLAVO'S GUIDANCE

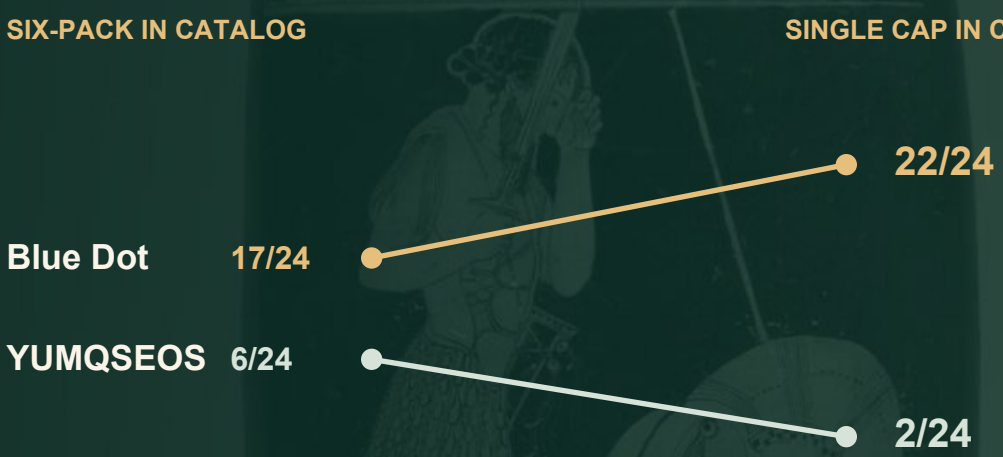
Investigate the complaint on the product itself. If it is real, address it before promising better performance in the listing.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

ASSORTMENT TEST · 23 OTHER OFFERS HELD FIXED

Neither new rival won. Blue Dot's picks still changed.

We replaced a $5.99 six-pack with a $5.99 single cap in the same slot. Neither entrant won a recommendation.



YUMQSEOS moved from 6/24 to 2/24. The swap changed the competing product as a whole, including its brand, pack size and reviews.

CONCLAVO'S GUIDANCE

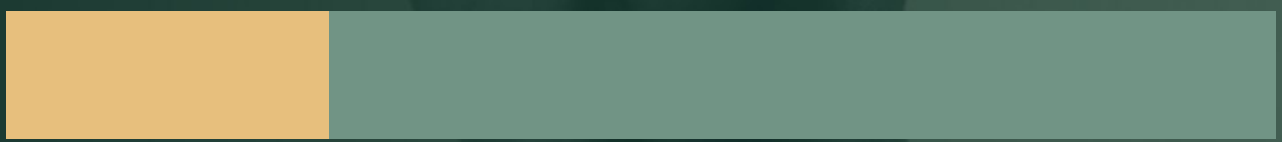
Repeat your test against a second realistic assortment before changing inventory or price.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

CONTROLLED REPLAY · 456 FRESH DECISIONS

The same evidence still produced different winners.

Each pair received the same buyer request, listings, reviews and photos in the same order. These cases came from earlier Blue Dot–YUMQSEOS disagreements.



58 pairs split

170 pairs agreed

In 25 direct switches between the two caps, both answers mentioned fit and water protection. One shopper could accept mixed fit reports at $4.08. Another could justify $8.90 for double-layer construction.

CONCLAVO'S GUIDANCE

Repeat both versions under identical conditions. A single favorable answer is common enough to mislead.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

SELLER TEST CARD

Decide what would make a change worth keeping.

STEP 1

Choose a buyer and a rival.

Write down the quantity, use case and product the buyer would otherwise choose.

STEP 2

Make one edit.

Compare the existing offer with one revised fact, price or pack. Keep the rivals and their order matched.

STEP 3

Score the final pick.

Count all trials. Track inspection and review openings separately. Repeat with fresh shoppers and a second assortment.

STEP 4

Check what happens in your store.

Set the smallest gain worth the cost in advance. Check purchases, returns and profit after variable costs with real customers.

CONCLAVO'S GUIDANCE

Use model tests to decide what deserves a live test. Use customer outcomes to decide what to keep.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

How the evidence was collected

Conclavo ran these private studies on November 27–28, 2026. One decision means a fresh shopper's proposed product or merchandise basket. In a paired comparison, the unit is two shoppers.

Study family	Fresh decisions
Two-pack choices in tape and batteries	48
Three battery pack sizes and small rerun	156
Unit-price confirmation and mixing-permission test	360 + 360
Battery ratings, lowest-price and best-value buyers	408 + 408
Four-cap study and wider search-first study	368 + 96
Four seller tests and sometimes-skipped listings	336 + 384
Fixed-evidence replay	456
Total	3,380

The 192-pair audit reuses the 384 sometimes-skipped decisions. Post-choice explanations and multiple comparisons of the same outcomes add no fresh shopping decisions.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Which packs filled the order?

Frozen U.S. Amazon prices. Each row pools two listing orders, two unit-price displays and repeat shoppers. A single or double contains one or two tape cartridges.

Minimum needed	Cheapest basket	Subtotal	Correct picks and totals
6 cartridges	Three doubles	$74.97	8/8
7 cartridges	Three doubles and one single	$91.45	8/8
8 cartridges	Four doubles	$99.96	8/8
84 batteries	Three 28-packs	$59.97	8/8
92 batteries	Three 28-packs and one 8-pack	$68.76	8/8
112 batteries	Four 28-packs	$79.96	8/8

All 16 baskets that saved money by including a smaller, higher-unit-price pack included it. All 32 where larger packs alone were cheapest omitted it. Each category scored 12/12 with unit prices shown and 12/12 hidden. All 48 advice choices carried through to correct proposed orders. At 92 batteries, adding the eight-pack saved $11.20.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Where did the first five-cent miss appear?

The three-pack search used 8-, 24- and 28-packs. Each minimum had all six listing orders under each price display. Counts below are the cheapest correct orders.

Minimum batteries	Unit price hidden	Unit price shown	Cheapest subtotal
8	6/6	6/6	$8.79
20	6/6	6/6	$19.94
28	6/6	6/6	$19.99
32	6/6	6/6	$28.73
40	6/6	4/6	$37.52
52	6/6	6/6	$39.93
60	6/6	6/6	$48.72
68	6/6	6/6	$57.51
76	6/6	6/6	$59.87
124	6/6	6/6	$97.49
199	6/6	6/6	$148.67
997	6/6	6/6	$719.54

Across the 12 quantities: 72/72 hidden and 70/72 shown. Both misses cost five cents extra. At 40 batteries a separate small rerun was 6/6 correct in both arms. Forty was the first sampled miss; larger orders were not systematically harder.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Showing unit prices did not repeat the earlier increase in mistakes

Two independent 360-decision tests used the same 40-battery requirement. Correct basket: two 8-packs plus one 24-pack for $37.52. The miss was 44 batteries at $37.57.

Listing order	Mistakes: hidden	Mistakes: shown	Correct: mix allowed	Correct: sentence removed
8, 24, 28	0/30	0/30	30/30	30/30
8, 28, 24	1/30	0/30	30/30	30/30
24, 8, 28	2/30	0/30	30/30	29/30
24, 28, 8	0/30	0/30	29/30	30/30
28, 8, 24	0/30	0/30	30/30	30/30
28, 24, 8	0/30	0/30	30/30	29/30

Confirmation: 3/180 mistakes hidden and 0/180 shown. Difference −1.67 percentage points, paired 95% interval −5.31 to +2.14; p = .25. The data did not reproduce the initial increase with visible unit prices.

Mixing: all 360 mixed packs. Cheapest basket: 179/180 with the permission sentence and 178/180 without. The buyer still asked which listing or listings to order. Of three misses, one shopper corrected itself afterward. Removing the sentence raised the error rate 0.56 points (95% interval −3.40 to +4.43; p = 1.00).

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Did ratings change the battery order?

Two separate batches varied the same 24-pack's stars and rating count independently. In each cell, 96 buyers saw a frozen set of competing prices. "Mention" comes from an explanation requested after the choice.

24-pack stars	Number of ratings	Lowest price: picks	Lowest price: mentions	Best value: picks	Best value: mentions
4.3	500	94/96	52/96	95/96	65/96
4.3	44,099	96/96	39/96	96/96	55/96
4.8	500	95/96	0/96	96/96	0/96
4.8	44,099	96/96	1/96	96/96	1/96

Across experimental cells, the cheapest basket won 381/384 lowest-price choices and 383/384 best-value choices. For the latter, raising stars or count moved picks by +0.52 points, simultaneous 95% interval −2.61 to +3.68. The buyer prompts ran in separate batches.

Original-rating and hidden-rating controls: 12/12 and 12/12 cheapest for lowest price; 11/12 and 12/12 for best value. Each group was small. Ratings were counts of stars. These cards supplied no written reviews. Under the lowest-price prompt, raising stars changed picks by +0.52 points (adjusted p = 1.00); raising rating count changed them by +1.56 (adjusted p = .50).

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Which changes moved Blue Dot's share?

The four-cap experiments tested private edits to genuine archived offers. Rows compare each setting averaged over the other changes in that experiment. Each denominator is the number of decisions at that setting.

Edit	Blue Dot picks	Change in points	Adjusted p[1]
Capacity language absent → present	1/48 → 48/48	+97.9 points	<.001
Retention claim absent → present	24/48 → 25/48	+2.1 points	1.00
Mean rating 4.1 → 4.7	56/64 → 64/64	+12.5 points	.023
Count of star ratings 100 → 1,200	59/64 → 61/64	+3.1 points	.688
Price $4.08 → $12.24	63/64 → 57/64	−9.4 points	.141
Displayed written-review count 30 → 500, total ratings fixed at 1,190	30/32 → 31/32	+3.1 points	1.00

Capacity and retention wording were varied independently. Stars, price and rating count were crossed in 128 trials. Written-review count was tested separately with actual review text held fixed. All 128 explanations referenced fit, 127 cost and 90 rating volume. [1] Smaller adjusted p-values give stronger evidence against chance differences within that experiment.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

How did price, stars and ratings work together?

Blue Dot picks with four cards already visible. Each cell had 16 shoppers. The other eight picks across all 128 choices went to SMPL objects.

Blue Dot price	Average stars	100 ratings	1,200 ratings
$4.08	4.1	15/16	16/16
$4.08	4.7	16/16	16/16
$12.24	4.1	12/16	13/16
$12.24	4.7	16/16	16/16

In a separate 64-choice browsing test, 4.1 versus 4.7 stars yielded 29/32 versus 32/32 picks (adjusted p = .50). With 30 versus 500 written reviews displayed, picks were 30/32 versus 31/32. The available texts and total star ratings stayed fixed.

Title wording	Without retention phrase	With retention phrase
Without capacity	0/24	1/24
With capacity	24/24	24/24

Without capacity language, the other 47 picks went to SMPL (40) or HAIRSOME (7). The $3.98 Deluxe cap won none.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

How much evidence did four-cap shoppers open?

Each ordinary or explicit-comparison arm below had 12 shoppers and the unedited catalog. “Alternatives” excludes the chosen product. Reviews were counted when their full text was returned.

Starting view	Other products inspected[1]	Distinct reviews[1]	Opened a photo[1]
Four visible cards	0 / 0	0 / 0	0/12 / 0/12
Cards plus browsing cue	3 / 3	24.9 / 30.8	2/12 / 3/12
Search first	3 / 3	25.2 / 30.5	10/12 / 8/12

[1] Ordinary / asked to compare. Products and reviews are means. Across all 152 first-screen trials, shoppers answered from the cards. In the original search-first groups all 24 chose Blue Dot, accessed all four products and requested critical reviews.

Six of 12 shoppers in each search-first arm went beyond a first review page. The median oldest review returned was 3,555 days old. No shopper sorted oldest first; 16/24 requested newest. All eight shoppers explicitly asked to open a photo did so. Interface comparisons came from separate batches.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

What changed with 24 products to choose from?

The two prompts saw identical catalog arrangements. Forty-eight fresh shoppers per arm. A card being shown, a product being inspected and a photo being opened are separate events.

Browsing outcome	Ordinary	Compare
Mean competing listings inspected, excluding winner	8.08	9.83
Median alternatives inspected (range)	8 (5–12)	10 (7–14)
Mean distinct search cards displayed	24.00	24.00
Shoppers who displayed all 24 cards	48	48
Shoppers who inspected all 24 products	0	0
Mean distinct review texts returned (total across trials)	34.35 (1649)	46.81 (2247)
Shoppers opening written reviews	48	48
Shoppers opening a photo gallery	11	5
Shoppers opening a reviewer JPEG	38	18
Distinct reviewer JPEGs per shopper, summed	53	27
Shoppers opening a gallery without opening an image	0	4

The comparison prompt added 1.75 inspected alternatives on average, paired 95% interval +1.02 to +2.50. Of 48 matched pairs, 31 inspected more, nine tied and eight inspected fewer. Photo opens fell despite more written reviews being returned.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Where did shoppers stop inspecting competitors?

Every shopper saw all 24 search cards. The distribution counts distinct competitors opened before the final choice. The chosen product is excluded.

Number of competitors inspected	Ordinary request	Compare request
5	6	0
6	1	0
7	12	4
8	10	5
9	8	12
10	6	11
11	4	10
12	1	3
13	0	2
14	0	1

The ordinary arm requested 625 review pages and opened 60 images including repeats. The comparison arm requested 876 review pages and opened 30 images. A gallery listing photo IDs counted separately from a request that returned a JPEG.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Which sellers earned a closer look?

Archived 24-product catalog. Every offer appeared in all 48 trials per prompt. Each entry gives two counts: ordinary request first, explicit comparison second. Sellers are listed alphabetically.

Seller	Inspected	Review text	Photo opened	Chosen
AKOAK	24 / 29	24 / 29	0 / 0	0 / 0
Ameliana	48 / 48	48 / 48	38 / 18	42 / 20
Auban	43 / 48	43 / 48	0 / 0	0 / 0
Betty Dain	44 / 42	44 / 42	0 / 0	0 / 3
CTEEAP	1 / 1	1 / 1	0 / 0	0 / 0
ELEBOX	48 / 48	48 / 48	0 / 0	0 / 0
EleCharm	2 / 5	2 / 5	0 / 0	0 / 0
Exceyes	48 / 48	48 / 48	0 / 0	0 / 0
Genergy	0 / 0	0 / 0	0 / 0	0 / 0
Geyoga	2 / 6	2 / 6	0 / 0	0 / 0
HAIRSOME	0 / 0	0 / 0	0 / 0	0 / 0
Hersome	0 / 3	0 / 3	0 / 0	0 / 0

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Which sellers earned a closer look? (continued)

Archived 24-product catalog. Every offer appeared in all 48 trials per prompt. Each entry gives two counts: ordinary request first, explicit comparison second. Sellers are listed alphabetically.

Seller	Inspected	Review text	Photo opened	Chosen
iSPECIAL	14 / 37	14 / 37	0 / 0	0 / 0
Jxuane	14 / 11	14 / 11	0 / 0	0 / 0
kasdfms	0 / 0	0 / 0	0 / 0	0 / 0
Kenllas	0 / 1	0 / 1	0 / 0	0 / 0
Kitsch	0 / 1	0 / 1	0 / 0	0 / 0
SMPL objects	32 / 48	32 / 48	0 / 0	0 / 0
Soft 'N Style	46 / 48	46 / 48	0 / 0	0 / 0
TONGTLETECH	2 / 4	2 / 4	0 / 0	0 / 0
Unaone	1 / 2	1 / 2	0 / 0	0 / 0
VEABEST	29 / 43	29 / 43	0 / 0	0 / 2
Yuest	0 / 2	0 / 2	0 / 0	0 / 0
YUMQSEOS	38 / 45	38 / 45	0 / 0	6 / 23

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

What happened when a fit complaint came first?

Two products were tested. A genuine contrary review was placed first or seventh in the default helpful sort. The same fit claim appeared in the details or the search card.

Claim location	Review rank	Inspected	Chose seller	Saw complaint	Repeats agreed
Detail	7th	24/24	9/24	0/24	9/12
Detail	1st	24/24	2/24	23/24	6/12
Search card	7th	24/24	11/24	1/24	10/12
Search card	1st	24/24	10/24	24/24	10/12

All 96 inspected the focal listing. Blue Dot wins rose from 11/24 to 21/24 with the card wording; Betty Dain remained at 0/24. With the complaint first, Blue Dot alone rose from 2/12 to 10/12; with it seventh, from 9/12 to 11/12. Small subgroups.

Across both sellers the card increased recommendations by 20.8 percentage points (95% interval +6.2 to +35.4). Moving the complaint first decreased them by 16.7 points (−29.2 to −6.2). The unaltered titles already contained some size language.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

What did a more descriptive photo label change?

One genuine Blue Dot customer photo was linked from the search card and its copy in the detail view. A descriptive label mentioned the reviewer's long braids. The image and product stayed fixed.

Label	Buyer	Opened linked photo	Opened any photo	Chose Blue Dot	Chose it without linked open	Repeat agreement
Generic	Ordinary	24/24	24/24	24/24	0/24	12/12
Generic	Compare	24/24	24/24	23/24	0/24	11/12
Descriptive	Ordinary	22/24	24/24	24/24	2/24	12/12
Descriptive	Compare	21/24	23/24	23/24	2/24	11/12

Linked opens: 48/48 generic versus 43/48 descriptive, a −10.4-point difference (95% interval −20.8 to −2.1). Blue Dot won 47/48 with either label. A link label itself conveys information; the comparison did not isolate the effect of adding the link.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Which competitors and positions changed the pick?

One test replaced an entire $5.99 entrant. Another moved Blue Dot deeper in the same 24-offer catalog and independently changed the number of results per page.

Entrant	Blue Dot	YUMQSEOS	Entrant	Other	Repeat agreement
TCOTBE six-pack	17/24	6/24	0/24	1	7/12
Betty Dain single	22/24	2/24	0/24	0	10/12

Blue Dot rank	Cards per page	Chose Blue Dot	Mean pages	Mean alternatives	Repeat agreement
7–12	2	23/24	12.00	7.33	11/12
7–12	6	19/24	4.00	7.67	7/12
19–24	2	24/24	12.00	7.25	12/12
19–24	6	22/24	4.00	7.75	10/12

All 96 shoppers in the page-size test reached and inspected Blue Dot. Six versus two cards per page changed Blue Dot picks by −12.5 points (95% interval −22.9 to −2.1). Later versus earlier rank: +8.3 points (0 to +16.7). The first six products were matched. In the separate entrant swap, Blue Dot gained 20.8 points (95% interval +8.3 to +33.3).

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Did moving fit wording help overlooked sellers?

Listings qualified if earlier shoppers sometimes skipped and sometimes chose them under the same buyer prompt. Each placement had 48 decisions per seller/prompt, balanced across 24 ranks.

Seller	Buyer	Earlier inspected; picked / 48	Inspected: detail → card	Picked: detail → card	Repeat agreement
VEABEST	Compare	43; 2	44/48 → 48/48	2/48 → 4/48	22/48
YUMQSEOS	Compare	45; 23	43/48 → 41/48	33/48 → 32/48	30/48
Betty Dain	Compare	42; 3	47/48 → 48/48	3/48 → 3/48	25/48
YUMQSEOS	Ordinary	38; 6	37/48 → 32/48	5/48 → 6/48	37/48

All 384 focal cards were shown. Inspections went from 171/192 to 169/192 (−1.0 points; 95% interval −6.2 to +4.2). Picks went from 43/192 to 45/192 (+1.0; −5.7 to +7.8). None earned more inspections and fewer picks. A second category was planned if inspections rose while picks fell for a seller or a pooled group. No such reversal occurred.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

How often did identical requests produce the same pick?

A repeat pair is two fresh shoppers given the same buyer request, rival order and offer details. They could still open different reviews and photos.

Experiment	Same winner	Different winners
Fit claim and negative review	35/48	13/48
Customer photo label	46/48	2/48
Replacement competitor	17/24	7/24
Page size and position	40/48	8/48
All four tests	138/168	30/168
Sometimes-skipped sellers	114/192	78/192

In the sometimes-skipped study, 40/192 pairs disagreed about whether to choose the seller whose wording changed. In 26/192 they disagreed about whether to inspect it. The next analysis traces the paths those shoppers took.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Did shoppers with different picks see different evidence?

These 192 pairs reuse the preceding 384 decisions. Each pair had identical prompts, catalogs and order. Shoppers chose different winners in 78 pairs; 57 of those splits were Blue Dot versus YUMQSEOS.

What matched?	Different winner	Same winner
Same 24 cards in the assigned arrangement	78/78	114/114
Same set of inspected listings	3/78	7/114
Same set of unique review texts	0/78	2/114
Same set of opened reviewer photos, including two empty sets	28/78	54/114
Same full set of tool response contents	0/78	0/114
Median number of review texts shared	34	32
Median share of distinct review texts shared, intersection divided by union	61%	63%

In 50/78 split pairs both inspected both eventual winners; 49/78 retrieved reviews for both. Three pairs received the same review texts for both winners, but all three saw other evidence differently. Of 156 shoppers in split pairs, 64 opened photos and four mentioned them. Different browsing alone does not explain the splits.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

Which products won with identical evidence?

Both original browsing paths from all 57 Blue Dot–YUMQSEOS splits yielded 114 packets, each containing the listings, reviews and photos one shopper had received. Four fresh shoppers per packet could choose any of the original 24 products. The source winner was hidden.

Product	Fresh picks	Source pick: Blue Dot	Source pick: YUMQSEOS
Ameliana Blue Dot	257/456	189/228	68/228
YUMQSEOS	154/456	22/228	132/228
Betty Dain	27/456	14/228	13/228
Auban	12/456	2/228	10/228
VEABEST	3/456	0/228	3/228
iSPECIAL	2/456	1/228	1/228
SMPL objects	1/456	0/228	1/228

Of 228 nonoverlapping pairs, 58 disagreed (25.4%; 95% interval 19.3–31.6%). Twenty-five switched directly between the two caps; 33 involved a third product. Ordinary requests split 5/44; comparison requests 53/184. Across four shoppers, 46/114 packets produced multiple winners. All six pairings per packet split 165/684 times; these pairings overlap.

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

APPENDIX / FULL RESULTS

What can a seller conclude from these tests?

The studies measure recommendations by one model, Codex gpt-6.1-sol at medium reasoning effort, under controlled requests. No accounts, orders or outside shoppers were involved in the Amazon trials.

CATALOG AND DATE

Battery and tape offers came from signed-out U.S. Amazon listings. Shower-cap records came from Amazon Reviews 2023 (All Beauty and Beauty and Personal Care). September 30, 2023 was the review-age reference. The historical offers do not establish today's price, availability or ranking.

CONTROLS AND UNCERTAINTY

Tests held rival offers fixed and varied selected facts in private copies. Listing positions were balanced or matched within comparisons. In the fit and photo tests, eight of 12 arrangements placed the focal item in the first half; matched treatments used the same order. Small groups and low baseline win counts limit precision.

WHAT A TOOL RECEIPT PROVES

A logged review or image reached the agent. Its private attention and reasoning are unobserved. Variant reviews can appear in a parent listing. The replay used previously contested cases and cannot estimate disagreement for all shopping requests.

WHAT REQUIRES FIELD VALIDATION

The experiments did not measure Amazon search, Rufus, human buying behavior, checkout, returns or profit. Merchandise subtotals omit shipping, tax and stock constraints. Test actual sales and contribution margin before scaling a change.

Public catalog source · Amazon Reviews 2023

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.

CONCLAVO

ORIGINAL RESEARCH



WORK WITH CONCLAVO

Which offer should you test next?

To sponsor a research report and/or request a client engagement, email contact@conclavo.com.

contact@conclavo.com

X @Conclavo_ Conclavo Strategy Room

Cover adapted from a photo by Jörg Bittner Unna CC BY 3.0

© 2026 Conclavo, Inc. All rights reserved. Please credit Conclavo, Inc. when sharing this report.