FEISTY AGENCY

The Testing Sheet

How to launch creative, read the winners, and not get fooled by your own dashboard

Most creative testing advice assumes the numbers on your screen are true.

They are not. Before any of the testing method in this document is worth running, you have to know how wrong your measurement is. So that comes first.

Part 1 — Your numbers are wrong. Here is roughly how wrong.

These are measured findings from a live Australian DTC account, audited in July 2026. They are not unusual. Assume something similar is true of yours until you have checked.

Finding 1 — The pixel reported 3.4× the real orders

Over a 14-day window, the Meta pixel fired **228 Purchase events** against **67 real paid orders** in the store ledger. The ratio held between 3.0× and 4.1× every single week.

Cause: the tracking plugin minted a *fresh random event ID on every load of the thank-you page.* Two loads of the same order receipt produced two different IDs, so the platform could not deduplicate a refresh, a back-navigation, or a reload. Its own "fire once per order" guard was a disabled checkbox in the free tier — it wrote the already-fired flag but never read it.

What this does to you: the optimisation algorithm was training on 3.4× phantom conversions. Every CPA figure in the account was wrong by an unknown factor, and — this is the part that matters — *wrong unevenly across ads*, because ads that drove more receipt reloads inflated more.

Check it yourself, today: count platform-reported purchases for the last 14 days. Count real paid orders in your store for the same 14 days. Divide. If it isn't close to 1.0, stop optimising and fix that first. Nothing downstream of a broken denominator is worth reading.

Finding 2 — The click ID was missing from 95% of orders

Share of paid orders carrying a Meta click ID (`fbc`):

Month	Share
April	23.6%
May	10.2%
June	14.7%
July	**5.1%** (7 of 136 orders)

The browser ID (`fbp`) was present on ~97% of orders, so the pixel was firing fine. It was specifically the *click stamp* that was absent.

Root cause, once found, was one line: the call-to-action component rebuilt its URL only for external links. Internal links froze the URL at the moment the component mounted — which was 2–4 seconds *before* the pixel finished writing its cookie. The link was built before the data existed.

The lesson generalises: attribution gaps are usually a race condition in your own front end, not a platform conspiracy. Look at your own code before you blame iOS.

Finding 3 — A "four-day sales drought" that never happened

The platform reported **zero purchases** across four consecutive days. The store took **three paid orders on every one of those days.**

Never read a platform zero-day as a sales drought. Reconcile against your store ledger before you change anything. Panic-editing a campaign on a reporting artefact is how a working account gets destroyed in an afternoon.

Finding 4 — A "17.9% payment rate" that was actually 100%

A checkout looked catastrophically broken: 54 orders started, 9 paid. On inspection, **45 of the 54 were internal test orders.** The real rate was 3 of 3.

Before you compute any rate from a store ledger, exclude your own test emails. Your agency domain, `example.com`, throwaway inboxes, and every order your own audit created. This takes thirty seconds and it has reversed conclusions.

The rule that comes out of all four

> ***When your ad platform and your store disagree, report it as a measurement gap — not as evidence that either side is lying.***

Store last-click attribution is broken in its own specific ways: it cannot see view-through, it cannot credit the ad → brand-search → purchase path, and it silently breaks across subdomains. On the account above, a checkout subdomain and the main domain were separate hosts, so the session cookie did not survive the crossing — which is what most "direct" revenue actually was.

So build your arguments on **attribution-independent facts**: within-store comparisons of orders actually created, or within-platform comparisons under one single event definition. Never one against the other.

Part 2 — How to structure a creative test

The finding that dictates the structure

On the same account, two ads ran at **0.12%** and **5.66%** link CTR.

They were **the same creative.** Same video, same copy, same hook. The variables were placement and audience.

47× spread, zero creative difference.

That single fact invalidates most of what is taught about creative testing. If you run four hooks across four ad sets, you have not tested four hooks. You have tested sixteen combinations with one observation each and you will confidently pick a winner that is noise.

The structure that follows from it

One ad set. All variants inside it. Nothing else changes.

```
Campaign:  [Objective] — Creative Test — [date]
└─ Ad set:  ONE audience, ONE placement set, ONE budget
   ├─ Ad A: hook 1  ┐
   ├─ Ad B: hook 2  │  identical everything else:
   ├─ Ad C: hook 3  │  same body copy, same CTA, same destination
   └─ Ad D: hook 4  ┘
```

- **Four variants maximum.** More than four and each gets too little data to separate.
- **Identical body copy and CTA.** If you vary two things you have learned nothing about

either.

- **Same landing page.** Obvious, routinely violated.

When to read it

Do not read before	Metric
1,000 impressions per ad	Link CTR, hold rate
50 link clicks per ad	Landing page bounce
15 conversions per ad	Cost per acquisition

Below those thresholds you are reading variance. The most expensive habit in paid social is killing an ad on day one at 300 impressions because it "looks bad".

The realistic benchmark

From the same measured account, ads with meaningful spend produced link CTRs spanning **0.12% to 5.66%**, with most between **2% and 5%**.

Use that as a sanity band, not a target. A 1.9% CTR on an ad that sells is better than a 5% CTR on one that doesn't, and this is not a rare case.

Part 3 — Reading fatigue

The signals, in the order they appear

1. **Frequency climbs past ~2.5** on a cold audience while reach stops growing. The ad is

now mostly re-serving to the same people.

1. **CTR falls while CPM holds steady.** This is genuine fatigue — the same delivery cost

is buying less attention.

1. **CPM rises while CTR holds.** This is *not* fatigue. This is auction competition or a

seasonal cost change. Do not refresh the creative; you will spend a filming day solving a problem your creative doesn't have.

1. **Conversion rate falls while CTR holds.** Not fatigue either. Look at the landing page,

the offer, stock, or your checkout — something downstream changed.

Only case 2 means "make new creative." The others mean something else, and the discipline of separating them is worth more than any hook file.

The refresh decision

Observation	Action
CTR down >30% from its own first week, frequency >2.5	New creative, same angle
CTR down, frequency <1.5	Audience problem, not creative
CPM up >40%, CTR flat	Wait. Nothing to fix.
Conversion rate down, CTR flat	Check the site, the offer, and stock — in that order

Part 4 — The trap that kills accounts

On the same account, **video creative consistently showed a worse cost per purchase than static creative.**

The obvious move is to shift budget out of video.

That is the wrong move. The video was doing the educating. The statics were closing people the video had already warmed. Turn the video off and the statics keep converting for about a fortnight on stored demand — then fall over. By which time you have lost the causal thread entirely and you are blaming the statics.

> ***Never rank two creative formats against each other on CPA when one feeds the other.***

Judge top-of-funnel on cost per qualified hold and on downstream volume. Judge bottom-of-funnel on CPA. They are a relay, not a race.

A related figure from the same audit, worth holding onto: at one point the reported cost per acquisition on a lane implied a spend equal to **124% of the entire store's revenue** for the period. The creative was not the cause. Acquisition saturation in a small national market was. **A CPA that has gone impossible is a market signal, not a creative brief.**

Part 5 — The sheet

Copy this. One row per test.

Field	Example
Test ID	`T-2026-08-01-hooks`
Hypothesis	"Contrast hooks beat POV hooks for cold AU broad"
Variable being tested	Hook line only
Held constant	Audience, placement, body copy, CTA, landing page, budget
Variants	4
Minimum data before reading	1,000 impr/ad
Primary metric	Link CTR
Guardrail metric	Cost per ATC
Store reconciliation done?	Y/N — **must be Y before any CPA conclusion**
Test emails excluded?	Y/N
Result	
Decision	Scale / iterate / kill
What I would not have known without this test	

That last row is the point of the whole document. If you cannot fill it in, you ran a campaign, not a test.

The five-minute weekly check

1. Platform-reported purchases ÷ real store orders. **Should be ~1.0.**
2. Share of orders carrying a click ID. **Should not be falling.**
3. Any zero-day on the platform — verify against the store before reacting.
4. Frequency on every cold ad set. **Flag anything over 2.5.**
5. One row added to the testing sheet. If none, you learned nothing this week.

Feisty Agency — feistyagency.com. Every figure in this document was measured on a live account, not estimated. The numbers are shared because most people optimising ads have never checked whether their denominator is real.