CONVERSION TEMPLATE

CRO Test Plan and Experiment Log

A hypothesis format that forces a real prediction, a sample size check that stops you calling winners early, and a log that keeps the losses.

Most conversion programmes fail on arithmetic rather than ideas. Tests get called at day four, wins get declared on 30 conversions, and the losing tests get quietly forgotten so the same idea comes back next quarter. This pack fixes all three.

WHAT'S INSIDE

START HERE

How to use this pack

Testing is a method for being wrong cheaply. That only works if you decide what would count as wrong before you start.

1. **Fix the leaks before testing.** A broken form or a five-second load time is not a test candidate, it is a repair.
2. **Calculate sample size first.** If you cannot reach it in a reasonable window, do not run the test. Ship the better-reasoned version instead.
3. **Write the prediction down.** A hypothesis without a predicted direction and size cannot be wrong, which makes it useless.
4. **Never call a test early.** The most common cause of imaginary wins is stopping the moment significance first appears.
5. **Log the losses.** They are more informative than the wins and they stop the same idea returning every six months.

THE HONEST BASELINE

Across mature programmes, roughly one test in five to one in seven produces a reliable win. If your log shows most tests winning, the tests are being called early or read wrongly rather than the ideas being unusually good.

SECTION 1

Before you test anything

Testing is the expensive way to find out something a five-minute check would have told you.

- [ ] Analytics is trustworthy and conversions are actually recording
- [ ] Every form submits and the notification reaches a human
- [ ] Checkout completes on mobile, on a real device
- [ ] Page loads acceptably on a mid-range phone on mobile data
- [ ] You have watched at least five session recordings of the page
- [ ] You have read what customers say in support tickets and sales calls
- [ ] The page states a price, or the reason it does not is deliberate
- [ ] The primary action is obvious without scrolling

Where the traffic and the money actually are

PAGE	SESSIONS	CONVERSION RATE	VALUE AT STAKE	TEST PRIORITY

TEST WHERE THE VOLUME IS

A 20 percent lift on a page with 200 visits a month is noise you will never measure. A 3 percent lift on the page with 40,000 is real money and reaches significance. Priority follows traffic multiplied by value, not by how much the page annoys you.

SECTION 2

The hypothesis format

FILL IN EVERY CLAUSE

Because we observed _____evidence_____, we believe that _____change_____ for _____audience_____ will cause _____predicted effect and direction_____. We will know this is true when _____primary metric_____ moves by at least _____minimum effect_____ over _____duration_____.

Why each clause is there

CLAUSE	WHAT IT PREVENTS
Because we observed	Testing opinions instead of evidence. Name the recording, the survey, or the number.
We believe that	Vague changes. One change, described precisely enough to build.
For	Testing everyone when the problem belongs to one segment.
Will cause	A prediction you cannot be wrong about.
We will know when	Moving the goalposts after seeing the data.
By at least	Declaring a 0.2 percent difference a win.

EVIDENCE, NOT IDEAS

'We think the button should be bigger' is not evidence. 'Nine of twelve session recordings show users scrolling past the CTA without pausing' is. The first clause is what separates a programme from a list of preferences.

SECTION 3

Sample size and duration

Do this arithmetic before building anything. It is what decides whether the test is worth running at all.

INPUT	VALUE
Current conversion rate on this page	
Smallest lift worth detecting	
Confidence level	95 percent
Statistical power	80 percent
Required conversions per variant	
Weekly sessions to this page	
Weeks required	

Rough guide to what is detectable

BASELINE RATE	TO DETECT A 10% RELATIVE LIFT	TO DETECT A 20% RELATIVE LIFT
1 percent	about 155,000 per variant	about 40,000 per variant
3 percent	about 51,000 per variant	about 13,000 per variant
5 percent	about 30,000 per variant	about 7,700 per variant
10 percent	about 14,000 per variant	about 3,600 per variant

READ THIS TABLE BEFORE PROMISING A TESTING PROGRAMME

These are approximate two-sided figures at 95 percent confidence and 80 percent power, and they are why most small sites cannot run meaningful A/B tests. Below roughly 1,000 conversions a month, you are better served by qualitative research and shipping well-reasoned improvements than by splitting traffic.

Run at least two full weeks regardless, so weekday and weekend behaviour are both represented.

SECTION 4

The test card

Test number : name Owner name

PAGE	START DATE	PLANNED END
URL		

FIELD	DETAIL
Hypothesis	
Evidence behind it	
Primary metric	
Guardrail metrics	
Minimum effect worth shipping	
Required sample per variant	
Traffic split	
Audience or segment	

What changes

CONTROL	VARIANT

GUARDRAILS MATTER AS MUCH AS THE PRIMARY METRIC

A change that lifts form submissions while halving qualified leads is a loss wearing a win's clothing. Name at least one downstream guardrail before starting.

SECTION 5

Reading the result honestly

- [ ] The test ran to the planned sample size, not to the first moment of significance
- [ ] It ran at least two full weeks, covering both weekday and weekend behaviour
- [ ] No changes were made to the page, the traffic mix, or the campaign during the test
- [ ] Guardrail metrics did not degrade
- [ ] The result holds on mobile and desktop separately, or the difference is understood
- [ ] Any segment claim was planned in advance, not found by slicing afterwards
- [ ] The confidence interval is reported, not just the point estimate

Verdicts

VERDICT	MEANS	DO
Ship	Beat the minimum effect, guardrails clean	Implement and note the date
Do not ship	No effect, or below the minimum	Log it and move on
Inconclusive	Ran out of time or traffic	Log honestly, do not rerun unchanged
Harmful	Variant lost	Log the insight, it is the most useful outcome

PEEKING IS THE MAIN SOURCE OF FALSE WINS

Checking a test daily and stopping when it first looks significant produces apparent winners at a far higher rate than the statistics assume. Decide the end condition up front and hold to it, or use a method designed for continuous monitoring.

SECTION 6

The experiment log

The most valuable document in a conversion programme, because it is the only thing that stops you retesting an idea you already disproved.

#	TEST	PAGE	RAN	RESULT	LIFT	SHIPPED	WHAT IT TAUGHT

Quarterly review

- [ ] Win rate is in a believable range rather than suspiciously high
- [ ] Shipped wins were re-measured a month later and held up
- [ ] Losing tests produced a written insight, not just a row
- [ ] Ideas are coming from evidence rather than from the loudest person
- [ ] The next quarter's tests target the highest traffic-times-value pages

RE-MEASURE THE WINS

Check shipped winners a month later. Some lifts are novelty effects that fade, and the only way to know is to look again once nobody is watching.

WHO MADE THIS

We build the systems behind the paperwork

Gatilab is an independent WordPress and search agency. We build sites, rank them, and stay long enough to prove the work moved something.

2014	850+	2,000+	10,000+
Founded	Clients served	Projects completed	Plugin installs

What we do

WordPress engineering

Custom themes, blocks, plugins, and migrations built to stay fast and maintainable after handoff.

SEO strategy

Technical fixes, topical coverage, and search visibility work measured against clicks, not scores.

Content marketing

Research-led articles and editorial systems your team can keep publishing without us.

Conversion design

Landing pages, service pages, and forms designed around the decision the reader is actually making.

Need this handled instead of templated?

Send the problem, not a brief. We will tell you whether it is worth building, what it depends on, and what we would deliberately skip. If the honest answer is that you do not need us, you will get that too.

Start a project

gatilab.com/connect

USE IT FREELY

Free to use, edit, and adapt for your own business and your own clients. No attribution required, no sign-up. The one thing we ask: do not resell it as your own product.

Gatilab

gatilab.com