

A/B-

SOCIVOLVE

TEST

by Coach Larisa Varma,
Certified Fitness Instructor

by
Nicolai Clöß
Socivolve

*PLAY
BOOK*

*Ab wann macht
Testing Sinn?*

Erst das Fundament, *dann* der Test.

Unterhalb einer bestimmten Traffic-Menge lassen sich keine Aussagen treffen bzw. Tests dauern einfach zu lang.

< 1000 Conversions / Monat

Kein A/B-Testing.
Erst Basics und Traffic Aufbauen.

1000 - 2000 / Monat

Grenzbereich: nur große Hebel,
frühe Funnel-Metriken.

2000+ / Monat

**Systematisches Testing, ab hier
lohnt es sich so richtig.**

Warum die Schwelle? Ein A/B-Test braucht eine Mindest-Stichprobe, um wirklich Aussagekraft zu bekommen. Liegt dein Traffic darunter, läuft jeder Test entweder monatelang, oder liefert Ergebnisse, die schlicht nicht stimmen. Deshalb gilt: Testen kannst du später immer noch.

Du liegst drunter? Dann ist nicht Testing dein Hebel, sondern die Basics:

- 1 Ladezeit, Mobile-UX, Trust-Elemente
- 2 Traffic aufbauen (Was für euch am Besten funktioniert: Meta, Google, Organic, whatever)
- 2 Qualitative Learnings ziehen: schau dir einfach mal 20 Session-Recordings an

Im Grenzbereich (1000 - 2000 Conversions / Monat)?

Teste nur große, deutlich sichtbare Änderungen und miss an frühen Funnel-Metriken wie der Add-To-Cart-Rate. Diese hat ca. 3-5x mehr Ereignisse als der Kauf, dein Test wird also schneller Aussagekräftig

Conversion Audit:
Daten statt
Bauchgefühl

Vier Quellen, *null* Vermutungen.

Der Audit gibt ein grundlegendes Verständnis für die Zielgruppe, Kaufmotive etc., aus denen alle späteren Hypothesen entstehen.

Vier Datenquellen gehören in jeden Audit:

1. Funnel-Analyse

2. Heatmaps & Session-Recordings

3. Qualitatives Feedback

4. Ist dein Shop auf deine Ads abgestimmt? – optional

1. Funnel-Analyse

Wo genau springen Besucher ab? Homepage? PLP? PDP? Checkout
Mobile und Desktop getrennt betrachten.

2. Heatmaps & Session-Recordings

Was machen die Leute, wenn sie in deinem Shop sind? Wie weit scrollen sie runter? wo springen sie ab? welche Buttons werden gedrückt?

3. Qualitatives Feedback

Umfragen, Support-Anfragen, Reviews. Hier bekommst du direktes Feedback, warum Kunden (nicht) kaufen, was die Probleme sind, was sie sich wünschen würden.

4. Ist dein Shop auf deine Ads abgestimmt?

Oft versprechen Ads das eine, der Shop das Andere. Wenn deine Kunden vor allem über Ads kommen, ist es wichtig deinen Shop auf diese abzustimmen.

*Hypothesen
entwickeln &
Backlog bilden*

Die *Wenn-Dann-Denn-Formel*.

Ein Test ohne Hypothese bringt niemandem etwas. Bevor du etwas änderst, mach dir Gedanken, was du damit erreichen willst.

Wenn

(kurze Beschreibung der Änderung)



Dann

(kurze Beschreibung des erwarteten Ergebnis)



Denn

(warum erwartest du das Ergebnis)

WENN der Hinweis zur kostenlosen Lieferung erst im Warenkorb statt bereits auf der PDP gezeigt wird, **DANN** steigen auf mobilen Endgeräten Conversion-Rate und ARPU, **DENN** mobile Impulskäufer werden nicht frühzeitig durch die Mindestbestellwert-Schwelle von 95 € abgeschreckt.

1 Änderung = 1 Hypothese

Jede Idee kommt ins Backlog

(Priorisierung klären wir im
nächsten Kapitel)

*Priorisieren:
Was wird zuerst
getestet?*

Framework: *PXL + 1*

Frameworks wie ICE fragen: „Wie hoch ist der Impact?“. Wüsstest du das, bräuchtest du keinen Test.

PXL stellt stattdessen objektive Ja/Nein-Fragen. Ich habe dieses Framework um eine Variable erweitert: *Wie schnell wird der Test signifikant?*

12 / 15

PXL-Score

Speichern

above_the_fold Is the change above the fold?	☒	/ 1
noticeable_5s Is the change noticeable within 5 seconds?	☒	/ 2
add_remove Does it add or remove an element?	☒	/ 2
high_traffic Does the test run on a high-traffic page?	☒	/ 1
mobile_relevant Does the change primarily affect the mobile experience?	☒	/ 1
user_testing Is it backed by a finding from user testing?	☐	/ 1
qual_feedback Is it backed by qualitative feedback (surveys, polls, interviews)?	☒	/ 1
heatmaps Is it backed by mouse tracking / heatmaps / eye tracking?	☐	/ 1
analytics Is it backed by a finding from web/digital analytics?	☒	/ 1
traffic_validity Can the page reach significance within 4 weeks of traffic?	☐	/ 1
ease_impl Ease of implementation (dev time): <4h=3, bis 8h=2, unter 2 Tagen=1, mehr=0	3	/ 3

PXL ist zum Anpassen gedacht. Wenn für deinen Shop ein Kriterium kritisch ist, nimm es als eigenen Punkt auf.
Wichtig ist nur: Die Fragen bleiben objektiv beantwortbar

*Testen:
Ganz oder
Garnicht*

Wie *lang* den Test laufen lassen?

Je kleiner der Effekt, den du erkennen willst, desto mehr Traffic brauchst du. Dabei gilt: halber Effekt $\approx$ vierfache Stichprobe.

Beispiel bei 2 % Baseline-CR:

+30 %  ca. 4k Besucher / Variante

+20 %  ca. 8k

+10 %  ca. 30k

Berechne die benötigte Sample-Size bitte immer vor dem Test. Es passiert leider immer wieder, dass Tools z.B. 90% Signifikanz anzeigen, die Stichprobe ist aber viel zu klein. Wenn jede Conversion das Testergebnis um 0,5% verändert, kann es nicht zuverlässig sein!

Geduld ist eine Tugend, oder so...

Lasse den Test mind. 2, max. 6 Wochen laufen.

Deine Änderung wirkt auf Weekend-Shopper anders als auf Leute die unter der Woche kaufen. Um diese Unterschiede zu berücksichtigen, sollten deine Tests mind. 14 Tage laufen.

50/50-Split + QA

Jeder Besucher wird zufällig zugeteilt und bleibt in seiner Variante. Vor dem Start: QA auf echten Geräten. Ein Layout-Bug in einer der Varianten führt zu Messfehlern und verfälscht das Ergebnis.

Nichts mittendrin ändern

Auch keine kleine Textänderung an Tag 4. Jede Änderung macht das Ergebnis unzuverlässig. Taucht ein echter Bug auf: Test stoppen, fixen, neu starten.

Mit welcher *Signifikanz* arbeiten?

Wer ist dieser Signifikanz

überhaupt? Sie sagt dir, wie sicher du sein kannst, dass der gemessene Unterschied wirklich existiert und nicht nur Zufall ist. 80% Signifikanz heißt: In 8 von 10 Fällen ist dein Gewinner ein echter Gewinner. In 2 von 10 war es statistisches Rauschen.

Sicherheit hat einen Preis. Von 80% auf 99% braucht derselbe Test 3,5x so viel Traffic. Statt 2 Wochen Laufzeit brauchst du 7. In dieser Zeit könntest du auch 2 neue Tests fahren, die evntl. einen deutlich größeren Impact haben werden.

Warum reichen 80%?

Tests pro Jahr = dein Traffic ÷ Stichprobe pro Test

Stichprobe bei 99% $\approx$ 3,5x Stichprobe bei 80%

⇒ Mit 80% fährst du 3,5x so viele Tests pro Jahr. Beim selben Traffic.

PS: Hier geht es um mind. 80% – Mehr geht immer

Selbst wenn man einen 80%-Test zweimal wiederholt, kostet er am Ende immernoch nur gut die Hälfte von einmal auf 99% warten.

Merke:

Solange der Fehler reversibel ist: 80% reichen

Business-kritisches (Preise, Checkout-Logik): teste auf 95-99%

*Learning: “Es
gibt keine
verlorenen
Tests”*

Gewinner, Verlierer, *egal*:

Der größte Denkfehler im A/B-Testing: Tests in Gewinner und „Verlierer“ unterteilen, und die Verlierer einfach vergessen.

Gewinner ✓

Fest einbauen => neue
Baseline, die bleibt

Verlierer ✗

Du weißt, was deine
Kunden NICHT überzeugt

Unentschieden

Die Änderung ist deinen
Kunden egal

Aus jedem Test, und vor allem, aus „Verlierern“ kann man lernen.

Jedes Ergebnis wandert ins Wissens-Archiv. Nach 10 Tests weißt du wie deine Kunden ticken und du lernst mit jedem weiteren Test dazu.

So sieht's in der Praxis aus:

Test 1 – Verlierer ✗

Die Preis-Schriftgröße auf PDPs verkleinern, damit der Preis “kleiner” wirkt. Ergebnis: negativ.

Das Learning: Kunden brauchen keine Tricks, sondern Klarheit.

Ein weniger sichtbarer Preis erzeugt Unsicherheit.

Test 2 – Gewinner ✓

Aus dem Learning kam eine neue Hypothese: Wer den Preis sofort sieht, entscheidet schneller. Also: Preis größer und prominenter präsentieren.

Ergebnis: ARPU +2,3%; AOV +2,1%;
Mobile ARPU +3,8%; **7.391€/Monat**

Zusammengefasst: aus einem als Verlierer abgestempelten Test hat sich eine neue Hypothese mit deutlich positivem Impact entwickelt.

Der nächste Schritt: ein kurzes Kennenlerngespräch.

Wir schauen in 20 Minuten, ob dein Shop und meine Arbeitsweise zusammenpassen. Wenn ja, analysiere ich deinen Store im Detail und präsentiere dir im nächsten Call 5 konkrete, priorisierte A/B-Test-Ideen.

- ✓ 100 % unverbindlich & kostenlos
- ✓ 5 Testideen mit Beobachtung und Hypothese
- ✓ Direkt umsetzbar – mit oder ohne mich



Kennenlerngespräch buchen – 5 kostenlose Testideen sichern