A/B testing

Background

A/B testing implies giving your users two **minimally different versions** (only one element/component), and then deciding which one fits their needs and goals better, based on analytics alone. **A/B testing is a purely quantitative insight**, and works best when used in conjunction with qualitative research methods. Not being a qualitative method, it is essential that it is carried out **with at least 30 users (ref @R)** per group. As with all quantitative methods, **the larger the sample the clearer the results.**

A/B Testing

100%
traffic → 50% | Version A | (Control)
→ 50% | Version B | (Challenger)
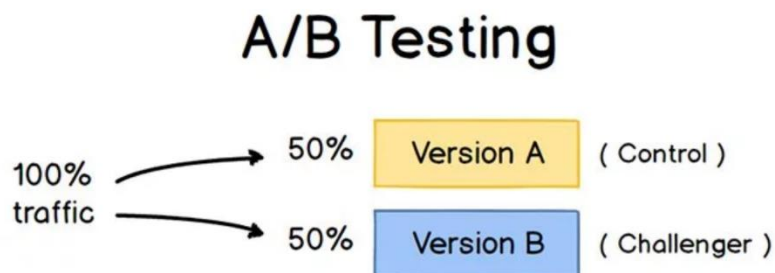
The two groups (control, challenger) need to be equally and randomly distributed in order to maintain the validity of the test. Variants used are created with key metrics (KPI) in mind, based on which user performance is assessed.

N.B. Testing more than one element/component (e.g. A/B/C test) is different. In this case, refer to Multivariate Analysis research methods instead.

Environment

Testing Method	Procedure
Split URL	Multiple URL leading to the different versions, live site, live users
Prototype	Carried out in-house, off-line, with recruited users
Multi-funnel	Rather than a page, this expands the focus on a particular flow that runs throughout multiple pages. **In short**: optimising with the end, not the very next step, in mind.

Methodology

7-step hypothesis

Because we observed (A)	**A = user behaviour**
and feedback (B)	**B = metric/feedback (DV)**
we believe that changing (C)	**C = layout/component/design/copy (DV)**
for visitors (D)	**D = segment of users (IV)**
will make (E) happen	**E = change we want to see**
We'll know this when we see (F)	**F = intended/expected metrics**
and obtain (G)	**G = goals/tasks/sub-tasks (IV)**

Research variables

Test variations

1. **Control** - represents the current version
2. **Challenger** - has one change from the current version, which can be a layout, component, design or content (copy) change

Types of metrics

Type	Examples
Sums & counts	number of users who visited page
Means, medians and percentiles	mean age of users who completed a form, mean time spent on a page, median latency of page load
Rates	number of click through pages
Probabilities	number of clicks with regards to a whole user-flow (funnel) & reaching end-goal completion, not just a step-by-step
Ratios	users who clicked on a CTA **divided by** total number of users

Choosing the right metric

- after clearly defining goals, tasks and sub-tasks
- after having a well-researched hypothesis based on which two variants will be tested (e.g. form completion/event registration <events>, heat maps <in-page navigation> - **see ECL examples below**)
- clearly defined metrics based on which goal completion on each variation can be assessed and compared (see defining metrics)

Events - analysis b/w components

A

🌐 **Jobs, Growth, Investment and Competitiveness**

Donec non elit bibendum lectus posuere auctor. Maecenas ac erat quam. Maecenas volutpat odio id maximus convallis. Cras nisi urna, blandit blandit efficitur id, euismod ac nulla. Ut rutrum tempus mi, quis lacinia eros. Donec bibendum vitae lectus vel tincidunt. Aliquam eget varius ex. Vestibulum tristique, massa et tempus lobortis.

🌐 **Jobs, Growth, Investment and Competitiveness**

🌐 **Jobs, Growth, Investment and Competitiveness**

B

| Item 1 | **Item 2** | Item 3 |

Donec non elit bibendum lectus posuere auctor. Maecenas ac erat quam. Maecenas volutpat odio id maximus convallis. Cras nisi urna, blandit blandit efficitur id, euismod ac nulla. Ut rutrum tempus mi, quis lacinia eros. Donec bibendum vitae lectus vel tincidunt. Aliquam eget varius ex. Vestibulum tristique, massa et tempus lobortis.

In-page navigation - analysis w/n components

A

PAGE CONTENTS
Policy areas ⌃

Policy areas

Latest

Background

Documents

lower prices for consumers. It enables citizens to travel, live, work or study wherever they wish.

B

PAGE CONTENTS
Policy areas ⌃

Policy areas

Latest

Background

Documents

lower prices for consumers. It enables citizens to travel, live, work or study wherever they wish.

Testing principles

- Hypothesis - **see 7-step hypothesis above**. Essential to have set-in-stone dependent and independent variables clearly defined
- Validity - **consistent behaviour over time** (takes into account learning curve, users successfully adapted to new user-flow with the changes made)
- Replicability - the results must be replicable longitudinally, not just a one off occurrence of chance
- Time-frame – time required to test for a satisfactory result
- User-base - self-explanatory
- Analysis & Choosing a variant - target behaviour change

Types of deliverables

- Heat-maps
- Click-maps
- Click through-probability-rates - assessing a whole user-flow (funnel) with the end-goal completion, not just a step-by-step
- User-flows
- Scroll-maps
- Content placing & ordering of content

Once you've picked your goal metric, think about how significant your results need to be to justify choosing one variation over another. Statistical significance is the most important part of the A/B testing process and the part that is often misunderstood. If the p-value is less than **0.05**, then there is a difference between the means and depending on whether that is a positive or negative, you'll need to identify and select the superior layout, component, design or content (copy).

Conclusion

Is it better than other research methods? Perhaps not, since all research methods are created equal.

However, with a statistically significant index offering practically significant results, A/B testing, used in conjunction with other qualitative methods, may provide more insights into what beneficial changes look like. The end-goal of A/B testing, like any other research method, is to provide better signposts on which better judgements can be made.

Frameworks

Free	URL
Wasabi (Open-source)	https://github.com/intuit/wasabi
Optimize	google link
Paid	
CrazyEgg	https://www.crazyegg.com
Optimizely	https://www.optimizely.com
FreshWorks	https://www.freshworks.com
Dynamic Yield	https://www.dynamicyield.com
VWO	https://vwo.com
HotJar	To be checked...
Adobe Target	https://www.adobe.com/marketing/target.html

Usage

Use it

- to assess in terms of broad, quantitative data the user preference on a layout, design or preferred component
- to compare two variations, such as improved versions (challenger) against older ones (control)
- to design small incremental changes (rather than compound)

Don't use it

- before pin-pointing goals precisely and the tasks the users must follow (see User-flows & Task analysis)
- if there are not clear indicators of the problem. This method cannot suggest what is missing or the expectations. In this case, it is best to assess qualitatively through interviews (see interviews) or user testing behavioural interactions (see user testing)

- to test new experiences or large-picture assumptions (exception sometimes when there are metrics to compare against)

- you have a compound solution or wildly different variations that need testing (A/B testing is most useful for small incremental changes)

Supporting references

[DataQuest - A/B testing definitive guide](#)

[Qualitative vs Quantitative (at least 30 users/test)](#)

[NNGroup](#)

[HubSpot checklist on A/B testing](#)

Example references

[Google - optimising copy through A/B testing](#)