



# Experimentation Checklist

Checklists ensure that there are minimum quality issues.

Here is an exhaustive checklist that the conversion rate optimizers and experimentation nerds at VWO created. The checklist is built based on 100+ experiments, incrementally and drastically improving our key business metrics. It will be useful for any experimentation team (from amateur to advanced) to avoid pitfalls and mistakes that come in the way of growth.

## **Observations stage** Make sure you tick at least one of the below

- Went through basic quantitative data available in tools like Google Analytics
- Did a thorough research on qualitative data using Funnels Analysis, Session Recordings, Heatmaps, Form Analytics
- Asked visitors about gaps between expectations and website experience through On-Page Surveys
- Conducted a Heuristic analysis to identify improvement opportunities
- Went through pages of competitors to see other ways of implementation

## **Planning stage**

- Converted the observation into a hypothesis
- The hypothesis includes a clearly defined and measurable metric which is known to have a business impact, directly or indirectly, along with secondary metrics
- Evaluated the hypothesis in comparison with other hypotheses on 3 parameters - Impact (I), Confidence (C), and Ease (E)

- Evaluation was not done in isolation but involved conversations and discussions with the CRO team or with at least 2 peers and definitely your manager
- Prioritized the hypotheses based on the ICE score with the highest scoring hypothesis taking the top priority
- Stages of the experiment are clearly known and configured in the program management tool used by your organization

## Launch stage

- Used one of the most robust and all-encompassing tools for testing, like VWO (which comes with a free hypothesis program management kanban board)
- Exercised due diligence to decide on the type of test:
  - Discrete changes in one single element of the page - A/B Test
  - Major redesign of a page leading to two or more pages - Split URL Test
  - Testing changes in two or more elements and how they interact with each other - Multivariate Test
  - Algorithmic or cross-platform tests or to minimize impact on the front-end due to client side changes - Server Side Test
- The metrics (primary and secondary) for the test are clearly defined in the hypothesis which is mapped to the test and configured accordingly in the test
- Appropriate targeting and segmentation is applied
- Guardrail metrics have been set so that no other critical business metric is compromised
- There are no spelling errors in the copy of the variations
- Images/GIFs included in the test are light in weight and suit the copy on the page
- URLs on which the experiment is being run have been verified
  - In case of split URL test use query parameter rel="canonical" in the variation URL to avoid duplicating content
- The test is previewed for rendering accuracy across multiple different browsers and devices (covering >80% traffic)

- Traffic split has been configured such that:
  - Sizeable proportion of the visitors will become a part of the test, while not compromising business continuity if something goes wrong
  - Both or all variations (and control) will get sizeable traffic to make the test worthwhile

[Use VWO A/B Test Duration Calculator](#)

- Tracking pixels are in place and are firing correctly, as per expectations
- Overlapping conditions have been verified such that the audience does not overlap across multiple experiments
- Integrations across different tools and platforms are working well

## Running stage

- You did not peek into the performance of the test
  - Peeking, without judgements or inferences is not a problem
- No metrics were changed during the test, especially the metrics included in the hypothesis
- The test was not paused during its run
  - If paused, all the data was flushed and the test was run again to ensure continuity and clean data
- The test was run for minimum 7 days to avoid weekly fluctuation effects and less than 3 months to handle for seasonality
- Tracking pixels have not gone through any changes in the way they were configured
- Ensure all variations are getting enough traffic and all goals are tracking and reflecting results.

If not, pause, flush data, re-evaluate, fix the issue and re-launch

## Conclusion stage

- Test is not concluded before 7 days and after 3 months
- Conclusion of test is based on probability of one version beating the other(s) (advanced) and/or statistical significance (basic)
- Evaluation of the test is done only on the metrics configured in the hypothesis
- Further deep dive is done into the performance of the variation(s) through heatmaps, session recordings, funnel analysis, form analytics, and other quantitative and qualitative secondary metrics which leads to more observations and ideas
- The best performing variation is shipped while other variation(s) are paused
- All the learnings (about the test, the audience, the systems, etc.) are drafted into a document
- Learnings doc and test conclusion is shared with the management and the larger team for alignment and sharing of knowledge which will be of benefit to business and team at large
- The next test is planned and charted out