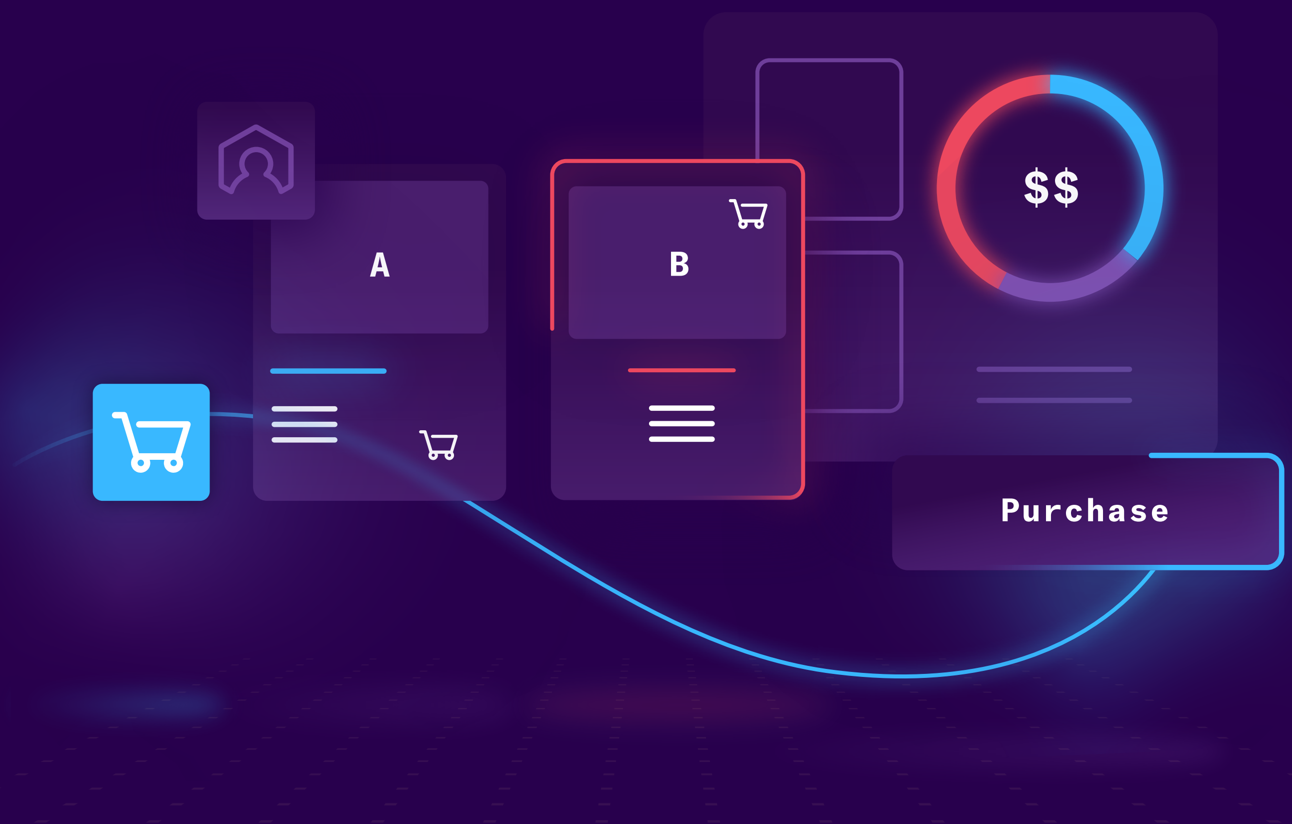


Heap

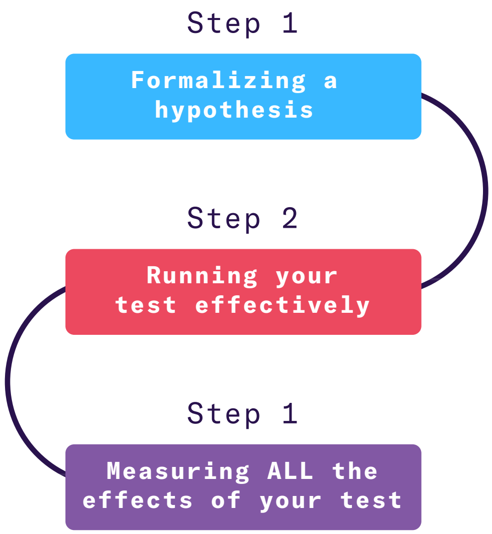
How To Do A/B Testing Right

A guide for Product, eComm, and anyone creating digital customer experiences

Introduction

A/B testing is a wonderful tool. There are few better techniques for optimizing your site and driving more conversions. We have the scientific method to thank for this: instead of changing things based on gut, or making decisions based on who is loudest or most senior, A/B testing—when done right—gives you verifiable, quantifiable information about how users interact with your site.

In our experience, however, lots of companies miss the opportunity to extract the full value from their A/B tests. Some teams waste energy by testing the wrong things. Some end up with misleading results because they don't follow best practices while testing. Others miss what is perhaps the most important element of A/B testing: ensuring your A/B tests produce positive results for the business as a whole.



Don't fall into these A/B test traps! In this paper, we tell you how to do A/B testing correctly by breaking up the process into three steps:

Step 1 - Formalizing a hypothesis

Step 2 - Running your test effectively

Step 3 - Measuring ALL the effects of your test

Read on as we go into more detail. Enjoy your A/B testing!

Step 1: Formalizing a hypothesis

At Heap, we believe that improving your product is a goal best reached when teams can be both creative and rigorous. This means two things: teams need the freedom to ask creative questions in order to form a hypothesis, and creativity is best served when ideas are tested through organized experimentation.

When it comes to A/B testing, many teams come up with a list of features, then start testing. That's fine, but they're missing a chance to be focused with their energy. That's why we recommend that every A/B test be run according to a **clear hypothesis** that articulates in advance what you're testing for and specifies why you think a change will make a difference.

Why is it important to do this work in advance? Four reasons:

1. It forces you to clarify the problem you're trying to solve, and makes you clarify how you're trying to solve it.
2. It forces you to articulate the change in behavior that will count as success.
3. It forces you to establish the KPIs that you're trying to move, and to spell out your mechanism for moving them.
4. Most importantly, it forces you to think in advance about downstream metrics, and how your A/B test may influence them. Yes, more people may choose one test option over the other, but what's really important is figuring out how whatever improvement(s) you make will affect the business. How does your test affect your bounce and exit rate? Or engagement? Or purchase? Or retention?

Below is a quick A/B Hypothesis Brief with key questions you want to answer before doing any testing.

What are we changing about the site/product?

What different behavior are we testing for?

What are the metrics that will tell us if our test is successful or not?

What will count as success?

What downstream metrics should we expect to change?

Here's what the Brief looks like filled out. Let's say your hypothesis is that changing copy placement on a page will cause more people to add a certain item to their cart. Here's how you'd answer the other questions:

What are we changing about the site/product?
We are moving the "ADD ITEM TO CART" button to the top of the page.

What different behavior are we testing for?
We think that people who see this button are more likely to add the item to their cart.

What are the metrics that will tell us if our test is successful or not?
Of the people who saw the button, more of them will have added the item to the cart.

What will count as success?
A 5% or more increase in the number of people who added the item to the cart, tested over 4 weeks.

What downstream metrics should we expect to change?
Track the number of people who added the item to their cart versus those who did not. Compare the average order value and repeat purchase rate between these groups.

Step 2. Running your test effectively

Ok, you've set out your A/B Hypothesis. Time to run your test. How do you do this right?

1. Choose the right tool

There are numerous tool available for A/B testing. Here are some of the best. They all integrate easily with Heap.

AB Tasty	LaunchDarkly	monetate A Kibo Company	Qubit.	ORACLE | maxymiser
split	Google Optimize	Optimizely	VWO	

2. Sample size and statistical significance

In A/B testing, if you don't run your test on enough people or don't test long enough, you won't get reliable results. Therefore, you need to calculate the right population and sample size, and make sure you run your test long enough to gather the right data. ("Population" means all the people who visit your site or app, and "sample size" means the people who will interact with the thing you're A/B testing.)

Since not everyone is a statistician, A/B testing tools such as Unbounce, AB Tasty, and Optimizely X offer calculators to help you figure out what sample size and number of testing days you need. You can also use Evan Miller's A/B Sample Size estimator.

The main goal of getting these numbers straight is to make sure your results are statistically significant. ("**Statistically significant**" typically means that there is 95% likelihood that whatever difference you measure isn't caused by random chance.)

If you've already run your A/B test, you can also check if your test reached statistical significance by using Visual Website Optimizer's statistical significance tool.

3. Run tests to completion

We've seen it many times: marketers start running a test, see some promising initial results, and stop immediately. After all, things look good - let's make that change and keep moving!

Here's the problem. If you haven't run your test to completion, it's likely you've stumbled onto a false positive. A "false positive" is a data error in which test results incorrectly suggest the presence of a condition which could simply be the result of randomness. In plain English, this means that **the less data you gather, the more likely the result you've seen will be due to randomness.**

If you prematurely stop the test early, or introduce new items that weren't a part of your original hypothesis, you won't gather enough data to know if what you're seeing is a true result, or a random fluctuation. Let your test run its course! Stay strong and wait for the results to come in. Then take action.

4. Learning from incremental growth

The more you make running A/B tests a habit, the more you're likely to find that many of them end up inconclusive. That's fine! **An inconclusive test is still a useful result** - it tells you that the thing you thought would make a difference (changing the color of a button, say) wasn't actually as important as you expected. Even that information can be enormously useful, since it tells you that your energy is perhaps better directed elsewhere.

You should also expect that many of your A/B tests will produce only incremental growth. This is as it should be! Incremental improvements, when implemented constantly and repeatedly, can pay enormous dividends. What's more important is that you're learning as much as you can from every experiment.

A few examples of web features to A/B test for more conversion or your desired behavior. Measure not just the raw numbers of people choosing each option, but the difference in revenue each group produces:

- Copy on any page
- Messaging on any page
- Location of buttons
- Your hero image(s)
- Recommendations
- Animations
- Product Descriptions
- Ratings
- Relative size of images
- Page length
- CTAs

Step 3: Measure the entire customer journey

As powerful as A/B testing is, many people overlook the most important reason for doing it: to optimize not just a local target, but to optimize the business!

The way to make sure you're doing this is to measure all the data the test produces. While many people use A/B tests to figure out whether people prefer option A over option B, the fact is the really useful information arrives when you start measuring the downstream effects of your change and knowing what those people did.

It's easy to lose sight of this. But what's important isn't quite whether more people click one CTA over another (though that is important); **it's whether more people bought products from you,** or if they bought more expensive products, or if they were more likely to return to your site, or if they were more likely to move off the trial plan to the paid plan. These are the metrics that really matter to your business - these are what your A/B tests are really aimed to improve.

So how do you make sure that your tests focus on these downstream events? Part of the answer was given in Step 1: when you make an A/B hypothesis, you think about what downstream events you want to measure, and you make sure to track them. The other part happens after the test, when you check days, weeks, and months later to measure the effects of your tests.

Here's an example:

For instance, you hypothesize that by changing the image of your product on your homepage, more new visitors to the site will click your free trial button. So you set up a test to measure this: half of the new visitors to your site will see the new image, and half will see the existing image.

After three weeks, you measure, and you find that 50% more of the people who saw your new image clicked the free trial button, compared to 39% of people who saw your old image. This is useful! Most teams would stop here, and decide to go with the new image. (Assuming they'd reached statistical significance.)

But the really useful information—the information that truly tells you what about your test was successful—arrives when you start measuring the downstream effects of your change.

For example, while 50% of people who saw your new image clicked the free trial button (that's 28% more than the percentage of people who saw the old image), how likely were those 50% to move to your paid plan? What kind of retention rates did they exhibit? How did they use your product in the trial stage?

If 11% more people clicked the free trial button, but that group of people were 75% less likely to move to the paid version, then in fact you've reduced overall revenue for your company. If they're less likely to keep using your product six months or a year from now, then you've actually increased churn!

You can run similar experiments with any A/B test.

The lesson here is don't just look at the immediate results of your A/B test to make a product decision; look at the downstream data to see how those results affected the entire customer journey. Measure as far down the funnel as possible and consider how it affects metrics like leads, click-through rates, visits, demo requests, and traffic-to-lead conversion rates.

Conclusion

Companies can build and improve their product most effectively when they take a data-driven, experimental, and iterative approach to A/B testing and product development. This involves forming and testing hypotheses, figuring out what to measure, making small improvements, and learning from every single experiment — whether successful or not.

At Heap we hope to help you produce maximum value with your A/B tests. Please get in touch for more information.

Have fun testing!