



# The Definitive Guide to **A/B TESTING** **DIGITAL PRODUCTS**

A practical guide to optimize your websites and mobile apps using a rapid, data-driven process

# CONTENTS

## INTRODUCTION

## WHY USE A/B TESTING

## A FRAMEWORK FOR A/B TESTING

- *step 1: generate a hypothesis*
- *step 2: track your goals*
- *step 3: implement your a/b test*
- *step 4: run your experiment*
- *step 5: deploy the winner*
- *step 6: rinse and repeat, always repeat*
- *case studies*

## A/B TESTING FOR RESULTS

## BEST PRACTICES

## DRIVING ORGANIZATIONAL CHANGE

## CONCLUSION



# INTRODUCTION

A/B testing is essentially applying the scientific method to business. You can A/B test everything from the subject lines of your marketing emails to the grade of lumber used in building houses.

This guide is specifically for product managers and product development teams that want to make better digital products using data-driven iteration. We'll give you a solid framework to approach A/B testing your digital products, best practices, and how to drive a sustainable cultural shift in your organization so that data-driven innovation becomes second nature.

***“The key to success is to **assume nothing and always test**. Start with the small things because there’s probably a lot of treasure right under your nose.”***

— Peter Gray  
Director of Optimization  
The Wall Street Journal



# WHY USE A/B TESTING?

You have no idea what your customers want until you actually give them something to try. It's great to interview your customers and conduct focus groups, but that usually just highlights problems with your product. A/B testing will pressure test your solutions and tell you what customers actually want.

A/B testing is also the only way to drive sustainable growth for your brand. Apptimize's internal data shows that companies who frequently A/B test are better at retaining customers than those who don't--so much better that companies who test grow their user base 4.2x faster than companies who don't.

*Companies who A/B test grow*

**4.2x**

*faster than companies who don't*

“

We A/B test everything. For example, messaging is known to have a lot of impact on user behavior. But A/B testing is how KAYAK decided to have orange buttons everywhere. For some reason, orange turned out to be the most profitable color.

—Vinayak Rinade, Director of Engineering



# A FRAMEWORK FOR A/B TESTING

## Step 0: Develop Your A/B Testing Infrastructure

Before you actually start your first A/B test, you need to develop the overall process. Here are few questions to answer:

- How will you actually implement the random allocation of users into different experiences? At a minimum, an A/B testing tool only needs to allow you to create more than one version of your website or mobile app, randomly allocate users to those versions, and help you track which users saw which versions. Will your team develop your own A/B testing solution or purchase a solution that is ready to use?
- How will the team generate A/B testing ideas, execute the tests, and analyze the results?
- How will A/B testing data be used in conjunction with other types of user data (e.g. general analytics data, user acquisition data, marketing engagement data, user testing results, etc.)?
- How will A/B testing fit into the existing product development process and roadmap?

The best way to answer these questions is to start with an A/A test. An A/A test is simply an A/B test where both versions of your A/B test are the same. It allows the team to run through an A/B testing dress rehearsal to work out all the mechanics of how and A/B test will run. It also gives the team a chance to review data and make sure that the results show no significant results.

Once you've completed your A/A test, you are ready to run your first A/B test.

## Step 1: Generate a Hypothesis

The first step to a successful A/B test is to generate a good hypothesis. The hypothesis will form the basis of your entire test, so it's important to pick it carefully. A test without a good hypothesis is like a road trip without a good map — you'll end up somewhere, but you won't understand where you are or how you got there.

A robust hypothesis will have three parts:

- 1) a result - the predicted outcome
- 2) a variable - an element of your app or site that can be modified
- 3) a rationale (why this will occur)

### A SIMPLE EXAMPLE

“If we move the sign up button to the top of the page, user registration will increase, because it will be easier for our users to start the process.” The signup button is the variable, the predicted increase in registration is the result, and the rationale is because it will be easier for users to do so.

#### DO MEASURE

- Click events like signups, link clicks, purchase completions, etc.
- Quantitative survey results
- Other quantitative data points such as minutes of video viewed, number of items added to cart, number of users invited to join, etc.
- Concrete engagement and retention metrics such as number of pages/screens viewed per sessions, session length, 1-, 7-, or 30-day retention, etc.

#### DON'T MEASURE

- Qualitative metrics like user sentiment: If you want to optimize for qualitative metrics, make sure to create qualitative approximations. For example, if you want to measure customer satisfaction, define quantitative survey KPIs such as Net Promotor Score.
- Vaguely defined customer engagement or retention: Make sure to create qualitative definitions for engagement and retention.

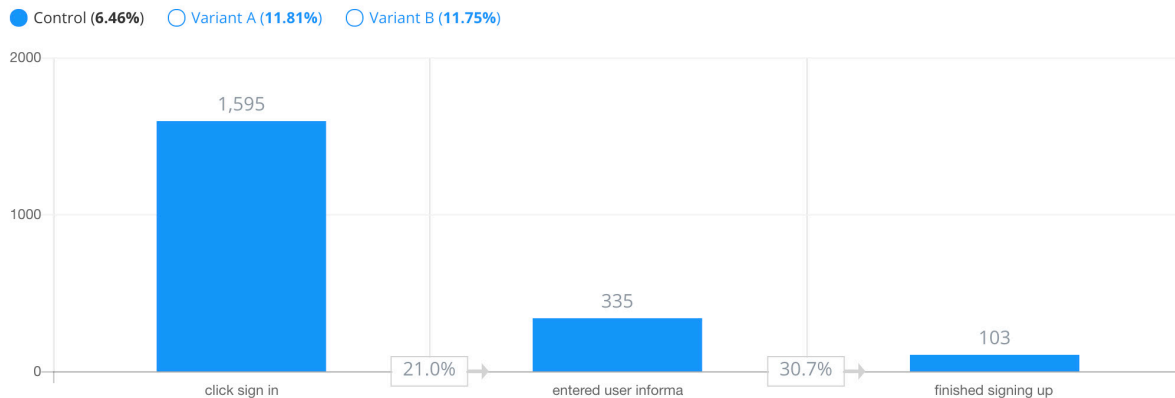
Strong hypotheses have results that are clearly measurable.

## Step 2: Track Your Goals

Once you have defined that results you are looking for, the next step is to track them. Most product teams will already have some analytics set up. When tracking data for A/B testing, make sure to not only track the main result, but also track the entire funnel.

### A SIMPLE EXAMPLE

If your hypothesis is “If we add more registration options, more users will sign up because they will find the option that is easiest for them,” then you will want to not only track the number of users who click one of the register buttons, but you’ll also want to track each step of the registration funnel.



In addition, you might also wish to track the number of registrations via each option plus those who bounce without registering and/or those who continue without registering.

### Step 3: Implement Your A/B Test

To do this, you'll have to isolate the variable you want to test, and then create a variant — make just one change to the variable like size, color or font. It's important to isolate and change only one element, so you can be sure that any changes in user behavior that occur are specifically because of that element. If you're not sure where to start, brainstorm elements that are most likely to influence the target metric.

### Step 4: Run your experiment

This part is self-explanatory: conduct the A/B test with your changed variables. Set the experiment live

## A CASE STUDY

**HYPOTHESIS:** Hotel Tonight, a leading hotel booking app that features at the last minute hotel deals, thought eliminating mandatory account creation during checkout will reduce friction, and thus increase bookings.

**GOAL:** Increase hotel bookings

**THE TEST:** The Hotel Tonight created two versions of the checkout flow—the original with a forced registration step and a variant B that removed the step and allowed users to book with just an email address for confirmation.

**RESULTS:** 15% increase in bookings

 Hotel Tonight





and wait for the results to come back in. Your app users or website visitors will be randomly assigned to either the original version or your experiment, and once enough users have engaged with your app or site, you'll be able to assess relevant metrics to understand how the experiment has impacted user behavior compared to your control group. Depending on the size of your user base, you may be able to run an experiment in just a few hours, or it may take weeks to collect a proper sample size.

## Step 5: Deploy the winner

Once you've got your data, you'll want to look for a result with 95 percent statistical significance and at least 80 percent statistical power. These numbers are industry standards that help confirm that these results are valid. If your results aren't statistically significant, consider running the experiment for longer until you hit that threshold.

Next, take a look at what this actually means. Was your hypothesis correct? If not, it's time to dig in a little deeper and try to figure out what happened. Experiments that show the variant to have a negative impact on a metric that you were looking to improve are still useful because it might demonstrate that your current version is better. Whether your hypothesis was proven true or not, you now have more information and data to build on for your next experiment. Take some time to be sure you really understand your results, so your next A/B experiment will benefit from your newly gained insights.

Finally, if there's a clear winner, it's time to figure out whether you want to deploy that version now or run an iterative test to build on it and generate more



uplift. If there's not a clear winner, use the new questions that you've created to design your next A/B experiment.

## Step 6: Rinse and Repeat, Always Repeat

Think of A/B testing as a cycle rather than a linear process.

Rather than a task that's done once and then abandoned, A/B experiments should be an ongoing driving force in your development. This cycle helps you gather meaningful data that can be used to improve nearly every aspect of your business. Be sure to document what worked and what could be done differently about each of your A/B tests to be sure that no insights are ever lost. A/B experiments should be the starting point for constant optimization, not the end.

## A CASE STUDY

**HYPOTHESIS:** The Wall Street Journal, one of the top US newspapers, believed that customers need different messaging in different contexts to encourage them to subscribe to the paper. The optimization team conducted a series of experiments to optimize the messaging within the mobile app. For this specific A/B test, they hypothesized that a simpler onboarding experience would eliminate noise and make it easier for customers to subscribe.

**GOAL:** Increase paid subscriptions.

**THE TEST:** They created a simplified version of the onboarding sequence which no longer included an introductory message from the publication's editor-in-chief and A/B tested it against the original onboarding flow.

**RESULTS:** That simple change led to a 23% uplift.

# THE WALL STREET JOURNAL.

# A/B TESTING FOR RESULTS

Now that you have an idea of what an A/B test looks like, what kind of results should you be looking for? Here are a few examples of the kinds of growth that successful A/B experimentation can have.

## INCREASING REVENUE

In essence, using A/B testing to increase revenue is simply finding the optimal way to get a user to buy something across your funnels. Test different algorithms that recommend purchases, suggest similar items once users have put something in a cart, or encourage users to add items into their cart. For streaming services, A/B experiments can offer an answer in regards to the best time to put up a paywall (will you retain users after giving them ten minutes of free content?). What's the most efficient way to get users down the funnel to the endpoint? A/B testing offers a multitude of opportunities to drive conversion and lift revenue.

## IMPROVE ONBOARDING

If users aren't getting value from your app right away, they'll be unlikely to return. A/B testing can help your team ensure users get value from the very first session. A robust onboarding process is an investment not only in growing a user base, but also in engagement and retention. This might take the form of a pop-up to welcome users to an app, a or a tutorial teaching users how to best use it or a multi-step process to personalize their app experience. An A/B experiment will help you understand how to best optimize your onboarding sequence for an enhanced user experience.

## INCREASE ENGAGEMENT/RETENTION

Once you've got users to download your app, you'll need to ensure they stick around and actually use it — otherwise all the effort to onboard them has been wasted. Examples of A/B experiments that can help increase engagement include trying different types of permissions to re-engage users (such as email, push notifications, or inputting a mobile number for texts) to see what works. For content-driven apps, look for ways to optimize your recommendations engine to encourage users to click on similar content to material they've already consumed. (Netflix, for instance, doesn't just serve up curated choices, but even displays custom artwork for their recommendations on a per-user basis, based on the user's past consumption preferences.) Experiments could also involve testing new features to add more value to the app and retain users over the long run.

## INCREASE VIRALITY

The virality of an app is really about the success in social sharing. A/B experiments can help you figure out the best way to get your users to share with a friend. Whether it's a promotion where inviting a friend gets both your user and their friend a financial reward, or a technique for reducing friction in the invitation process (i.e., adding a simple "send to a friend" button), increasing virality is one important aspect to

growing and maintaining a robust user base.

### **REDUCE COST**

Another key part of optimizing your app is reducing the cost of overall business operations. Particularly in industries like telecommunication or banking, each time a user turns to the app instead of calling the business saves money. A/B testing can help find the best channels to encourage users to use the app as a support center rather than calling. With this kind of testing, it's easy to connect the pipes of data to check if the changes from A/B experiments are actually successful in reducing calls while delivering on customer satisfaction.

### **GROWTH**

Many businesses think about growth in linear terms: paid advertising leading to a larger user base. While you can use A/B testing to optimize ad copy and creative to see what makes your dollars go farthest, that's really only half the battle. It's essential to invest in maintaining and keeping these users engaged, by building a "sticky" process that ensures you are gaining long-term and loyal users. Rigorous A/B testing can ensure your advertising dollars are having the largest possible impact, by optimizing processes to continually engage your user base and lower attrition.

## **BEST PRACTICES**

Here are some elements to keep in mind for A/B testing success.

### **ALWAYS HAVE A HYPOTHESIS**

Without a good hypothesis behind your A/B experiment, you're driving without a roadmap and that means you can miss out on a lot of valuable insights. Even if you get a clear result, you'll lack some important context for understanding why that occurred. In order to get the most benefit, you want a carefully crafted hypothesis that is both specific and testable. The best hypothesis is one that's informed by research. Whether it's customer feedback, or similarities to a previous test, data helps make a hypothesis strong. The more specific a hypothesis is, the more detailed your conclusions can be. It doesn't matter as much whether your hypothesis is proven correct or not — a good hypothesis will give you actionable data whatever the outcome is.

### **HAVE ENOUGH USERS TO REACH A CONCLUSION**

Obviously, the more users you can include in testing, the better results you can get. However sometimes huge sample sizes just aren't feasible. In a nutshell, the larger the difference between variants, the more confident you can be that the results are statistically significant with smaller samples. In order to calcu-





late this exactly, you'll need to know how big the difference between variants will be, how many variants there are, and what the conversion rates are.

### **ALLOW ENOUGH TIME**

It's actually possible to calculate how long you'll need to run your test in order to see conclusive and significant results. Multiply the fraction (percentage) of users in the test with the amount of monthly users divided by four (for weeks). Then multiply the number of variants by the number of users needed per variant to be statistically significant and divide this number by the first result you calculated. The resulting sum will be the number of weeks needed for the test, allowing you to plan accordingly.

### **LIMIT CROSS-TEST CONTAMINATION**

Be aware that there are some tests that shouldn't be run at the same time, since they can interfere with the others' data. Tests that interact strongly should either be run separately and at different times, or as a multivariate test. The multivariate option will make sense if the tests both optimize the same goal and run in the same page — in other words, if they're compatible.

### **DON'T TEST TOO MUCH AT ONCE**

Put simply, don't bite off more than you can chew. Testing too many variants is not a good idea for a few reasons. It can be overwhelming for you team, and could interfere statistically with the validity of your conclusions. The industry standard for a confidence level is 95 percent, meaning that there's a 5 percent chance of a false positive. When you introduce multiple variants, that error probability increases exponentially. Keep your variants well-conceived, data and hypothesis-driven, and as low as possible. About three variants plus the control is about as high as you should go.

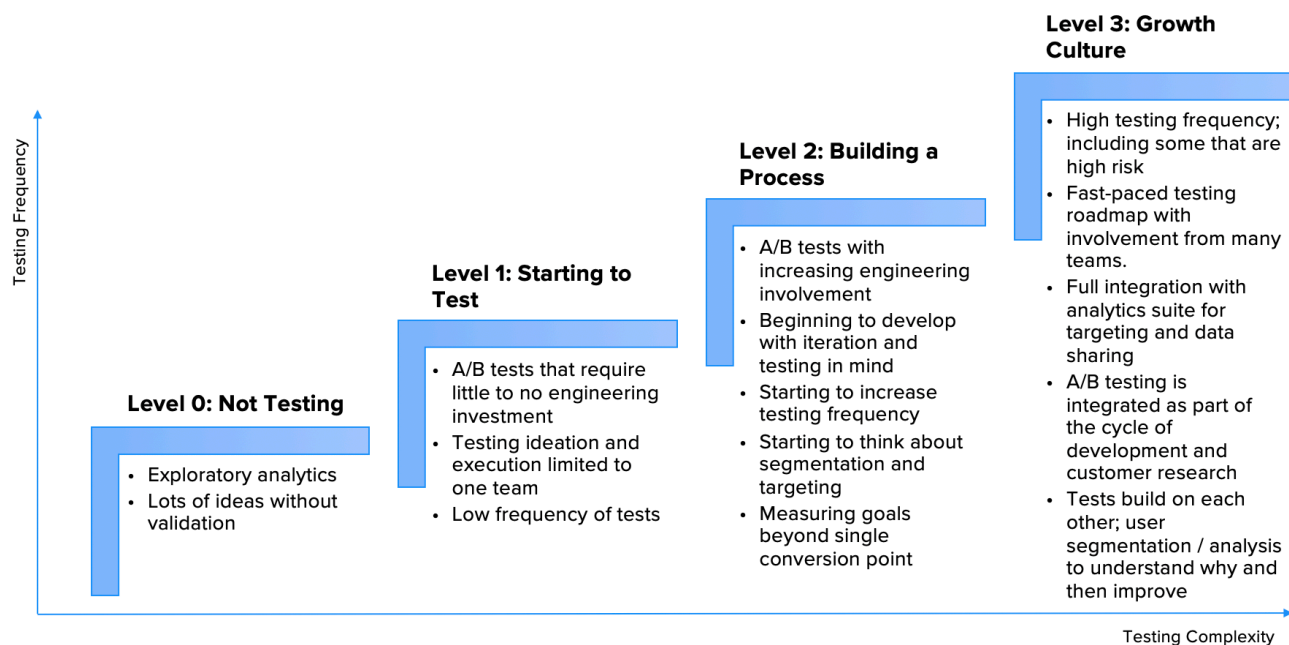
# DRIVING ORGANIZATION CHANGE

One of the hardest parts about A/B testing isn't the testing itself—it's making sure that everyone involved in your organization is actually buying into the idea of constantly challenging themselves and doing the hard work of testing their assumptions. For A/B testing to be a success, you'll need everyone — your product, development, and marketing teams, analysts, and executives — all on board. It sounds like a tall order, but it's possible. Here are some tips that have worked for companies that have successfully implemented a mindset shift in the organization.

## USE EARLY WINS TO PROVE OUT THE VALUE

Don't just dive into the deep end. It's a good idea to run small tests with relatively low stakes that prove out the concept, such as visual tests to assess how different creative impacts conversion rates. Once you have these early wins, circulate them with your stakeholders and teams.

As you gain traction across the team (and stakeholders in other teams), A/B testing velocity will increase. Ultimately, the goal is to guild a culture of iteration within the company that generates sustainable, predictable results that build on each other (see the figure below).



## **BRING IN STAKEHOLDERS & ALIGN TO THEIR PRIORITIES**

The next step in this process is generating buy-in from these stakeholders. The best way to approach this is to identify what these stakeholders care about. What do the developers need? What do the executives want? What will marketing find valuable? Bring all your different departments into the fold by aligning A/B testing with the priorities they care about. This is where contextualizing the early, small A/B tests for each department will pay off. By tying A/B testing to metrics and goals each of your departments care about, you'll have an easier time getting everyone on board.

User testing, including A/B experiments, is an investment. Generating buy-in as early as possible will help you see returns on this investment as quickly as possible. Emphasize that this is a team effort that will bring rewards to everyone, not just one department over another. Product teams get data-driven usability insights, developers save time by reducing the time spent rebuilding features, marketers get new selling points, and executives get insight into new ideas. Point out that everyone gets something valuable from A/B testing. After all, it's possible that sometimes the results of an A/B test won't make a person or a department happy. When you generate this kind of buy-in at the outset, it's more likely that the departments and individuals will be more willing to accept the conclusions and act on them. This will save you valuable time in the long run.

## **A/B TEST THE RIGHT THINGS**

Another way to be sure your teams adopt the cultural change needed for data-driven sustainable growth is to be sure that what you're testing actually matters. Identify the core goals of your business, and be sure that each test you run not only aligns with these goals, but moves your company closer to actually achieving those goals. Be sure that you are testing the elements that drive the KPIs your whole organization cares about. This dedication is how you become a more modern digital team.



# CONCLUSION

In a quickly evolving marketing landscape with growing market pressures, new devices, and ever-changing user behavior, the need for an iterative approach to product development and innovation using data-driven insights is more important than ever. It's almost impossible to predict what will resonate with your customers, thanks to constantly shifting behaviors and preferences.

Game changing innovation just isn't possible without an experimental mindset focused on continual optimization and improvement. A/B testing offers that iterative approach to innovation, along with incredible insights into user behavior that can be leveraged into greater conversions and productive, sustainable growth.

A good A/B experiment will not only give developers information on how the goal performed (and thus whether or not to implement it) but it should also generate more questions. Why did this goal perform differently? Can we improve this performance? What does this result tell us about our user base? While A/B testing is valuable for the data it directly yields, what users choose to do with that data and how they are able to build upon it is where the true value lies. Successfully taking lessons from these experiments is the key to innovation. And although A/B testing can be hard work, it's the only true way to sustainably drive innovation for modern organizations.

“

There was one really minor change that we made.... I was shocked by it! My gut said no... but I was proven wrong! I was proven wrong by data.”

—Corinne Almirol, Product Manager

**vevo**

# ABOUT APPTIMIZE

Apptimize is an innovation engine that provides A/B testing and feature release management for native mobile, web, mobile web, hybrid mobile, OTT, and server. Industry leaders like Hotel Tonight, Rakuten, and Glassdoor have created amazing user experiences with Apptimize.

LEARN MORE AT [APPTIMIZE.COM](https://apptimize.com).