

A/B Testing



With A/B Testing you can split traffic randomly into different groups and show each group variations of a message. This is done by creating multiple *experiences*, then assigning each a percentage of traffic.

This Article Explains

This article details how to use A/B Testing to split traffic randomly into different groups and show each group variations of a message.

Sections in this Article

- [What is an A/B Test?](#)
- [Create an A/B Test](#)
- [Things to Consider Prior to Testing](#)

What is an A/B Test?

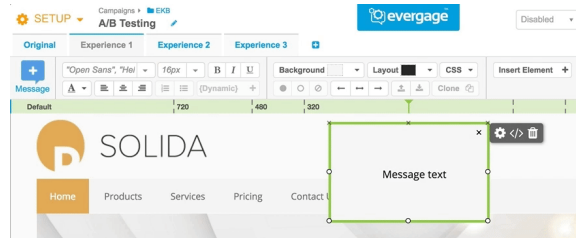
A/B testing, in its true definition, is a form of statistical hypothesis testing with two variants: test and control. Within Evergage, we use A/B testing to more broadly define any campaign that uses a randomized split in the traffic to test a hypothesis. There are several things you should understand before creating an A/B test:

- When creating an Evergage campaign, you can create multiple experiences which are shown to a set percentage of traffic as determined by your needs
- As a best practice for an unbiased, statistically sound test, create messages as variations of each other. For example, testing a popup against an inline message is not a true measure of effectiveness as both invoke different behaviors for the users and cannot be compared. A better example of an A/B test would be a campaign that tests which exit popup discount code is more effective

For a better understanding of how you can design a campaign such that the results are statistically sound and do not contain any biases, please contact your Customer Success representative for guidance.

Create an A/B Test

1. [Create or Edit a Web Campaign](#)
2. Create the necessary experiences and messages
3. Click **SETUP**
4. Select **EXPERIENCES**
5. *Test Mode* is set to **A/B** by default
6. Use the sliders or enter a percentage for each experience



NOTE

Changing the percentage of traffic on one experience adjusts percentage (s) on the remaining experience(s) to total 100%. If you adjust user percentages after publishing your campaign, visitors who saw the campaign prior to the change will continue to see the experience (or control) they saw first.

7. If needed, set the **Control** percentage of viewers who will see the original page experience

8. Click **SAVE** or **SAVE & CLOSE**

Things to Consider Prior to Testing

Give the Test Adequate Time

- **Resist a rush to judgement.** In the early days of test, it is best to ignore the campaign statistics since they can change before stabilizing. This is especially true for low-traffic tests.
- **Use multiples of a full week of data.** When you are deciding at test, a best practice is to include multiples of a full week of data, since behavior will vary across days. You can set the beginning and end dates on the *Campaign Statistics* screen.

The End of the Test is Not the End of the Story

- **Post test lift may be higher or lower.** To set expectations, even if you have a winner, you may not replicate the exact lift you saw in the experiment. The key is that the winner's metric will be higher than the loser's.
- **Check for "Shiny Object Syndrome."** When you do see a winner, a best practice is to check whether you see the same effect across first-time and returning visitors. If you don't see the effect for first-time visitors, this lift may be due to novelty or "shiny object syndrome." Your regular visitors are curious about the change, but will eventually revert to their previous behavior. If you do see the effect for first-time visitors, it is more likely to be lasting.

Decide Whether to Use a Sample Size Calculator

- **How long you run a test should be driven by your data rate and the size of the effect you're looking for.** You may want to consider using a sample size calculator when planning your test. There are many available online including: <http://www.evanmiller.org/ab-testing/sample-size.html>. This will give you an idea of how much data you need for a given test, and that in turn will give you a sense of how long to run the experiment. In classical statistics, you do this in advance, and check your result once (and only once) when you have the required data; at that time you declare whether any difference you see is real or due to chance.
- **Using a sample size calculator is optional.** Sample size calculation is known as a power analysis, and is a required step in classical statistical testing. Because Evergage uses a Bayesian approach rather than a classical framework it is not required that you calculate your sample size, but it may help you with planning if you're unsure about how long a test should run.
- **Compute the sample size and run the test.** If you decide to use a sample size calculator, if you have reached the sample size and have not reached 95% confidence, you can stop the test and declare there was no difference. (You can use a difference confidence if you like.) If you have reached 95%, you can declare a winner.
- **You may reach confidence sooner than expected.** In contrast to classical testing, Evergage models the underlying distributions directly, and typically requires fewer samples. So, if you are running your test for a month, but see confidence at 3 weeks, you can stop and declare a difference.