

webtrends®



webtrends®

Website Testing 101:

Understand the Differences
Between A/B and Multivariate



White
Paper



Overview

Testing has long been used in direct mail and print advertising to determine which elements of a campaign receive the better response. The same is true in testing elements on a website or landing page. While every optimization opportunity is different, it is important to know the types of tests that are available and how they are used to help you create a website that inspires conversions and can improve your ROI. There are two main types of tests: the A/B and the multivariate. Understanding the differences between A/B and multivariate testing helps marketers select the proper method for reaching optimization goals and improving customer experience.

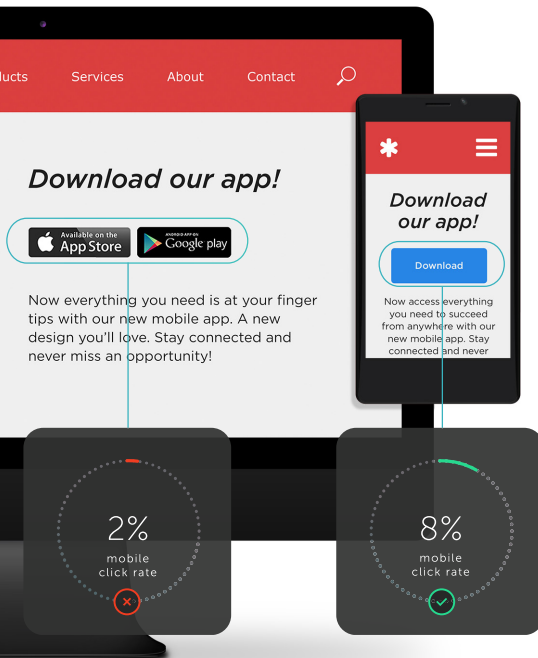
A/B testing

The most common starting point for optimizing a website is the A/B test. By using this testing method, one design or group of elements or copy is compared to another distinct design or group. On a landing page, an A/B split test is the competition of two distinct pages, where a portion of live traffic, usually 50 percent, is sent to one page and the rest to the other. The winner of the A/B test is the page that provides the highest conversion rate, or the key performance indicator that is most useful to the marketing campaign or organization.

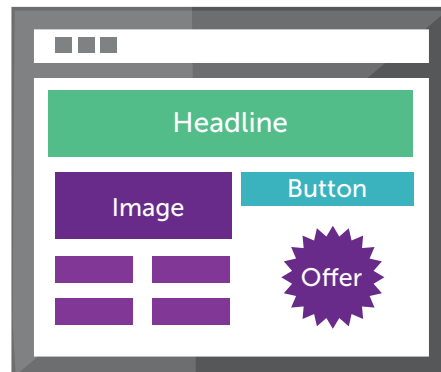
There are three types of A/B tests all of which should be considered when beginning a website optimization campaign: the template test, the new concept test and the funnel test.

Template test

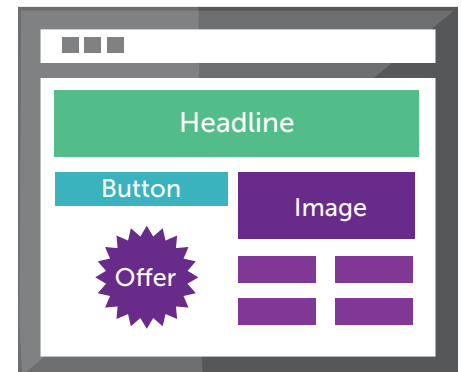
In this test, the same general content – headline, image, offer and button color – is compared with a different layout and/or creative treatment. The goal of this test is to discover which placement of the content better emphasizes the value proposition and improves readability. It often sets the groundwork for a multivariate test.



Use the template test when you want to ensure you have a solid design, before or after testing messaging. It is most often the first test to conduct.



Test A

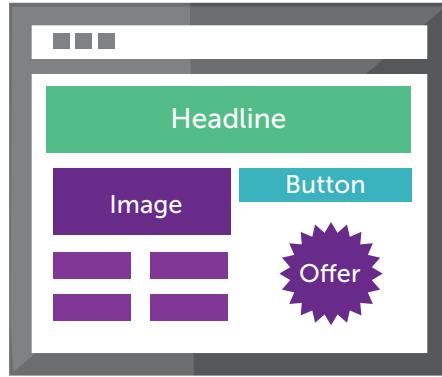


Test B

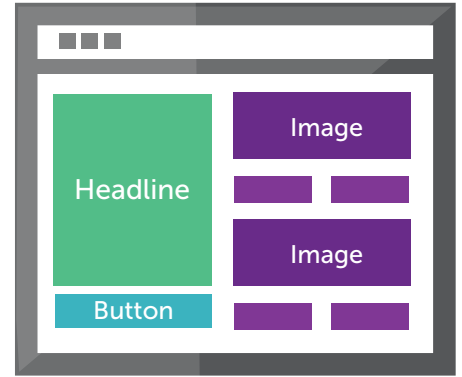
New concept test

This test compares very distinct pages, a current or control page with one that offers a totally new approach. Sometimes this involves introducing new functionality, animation, interactivity or other dramatic steps. However, it can also be on a smaller scale, such as introducing new messaging or offer that requires a complete redesign.

Use the new concept test when your current page has already been tested many times and conversion rate increase has stalled, or you believe the way to really engage your visitors is through a noticeable change.



Test A

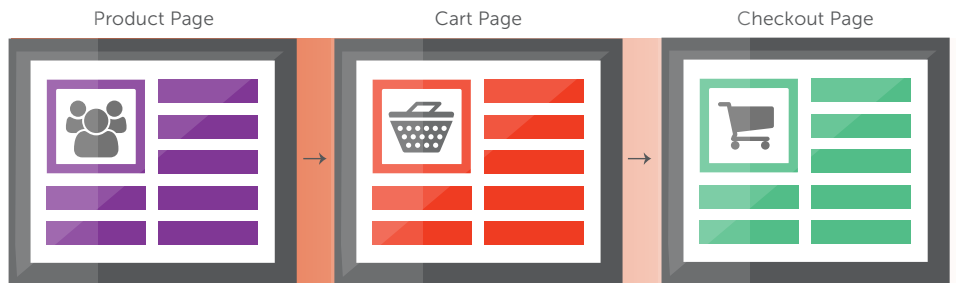


Test B

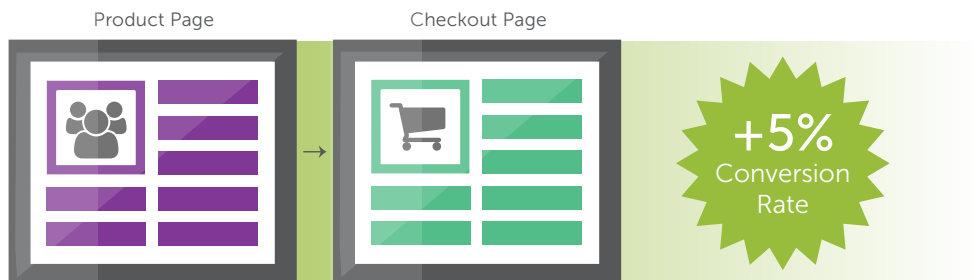
Funnel test

This A/B test compares different multi-page experiences in order to see which one results in better conversions. For instance, you can send users to a no registration page or to one requiring registration, or a one-page form vs. a three-page form.

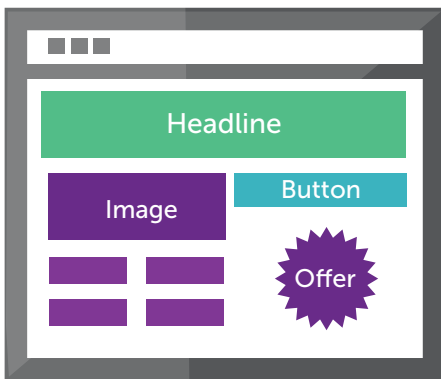
Use the funnel test when you want to test content that extends past one page. While a funnel test can be more technically demanding, it should be done early in the testing process so that you don't optimize a page and then find out it's actually a suboptimal experience.



Test A



Test B



This wireframe example represents a common landing page layout. To properly design a multivariate test, it requires you to understand page elements within the framework of three key terms: factor, level and experiment.

Multivariate testing

While multivariate testing is less commonly used, it can often produce better results than A/B tests, particularly where there is complexity. The technical and statistical aspects of multivariate testing can seem complicated, but understanding the basics of what constitutes a successfully designed test is important before undertaking one.

Framework of three key terms: factor, level and experiment

1. Factor

An element of the web page (headline, image, offer, button, etc.) being tested. The element can also be groups of content, e.g. left column, button and image together, or all banner ads on a page.

2. Level

This is content that is assigned to a specific factor to be tested. In the illustration, there are four factors taken from the example page (headline, image shot, offer and button). Each of those four factors is assigned four levels. Note that the levels of one factor do not have to relate in any way to the levels of other factors.

3. Experiment

This makes use of both factor and level in unique combinations.

In a multivariate test, experiments are shown randomly to live visitor traffic. Each experiment is shown to many people, but each person only sees one experiment. Once a statistically significant number of conversions are reached, analysis determines the elements on a page that create the highest conversion rate. The optimal combination is the one experiment that performs the best.

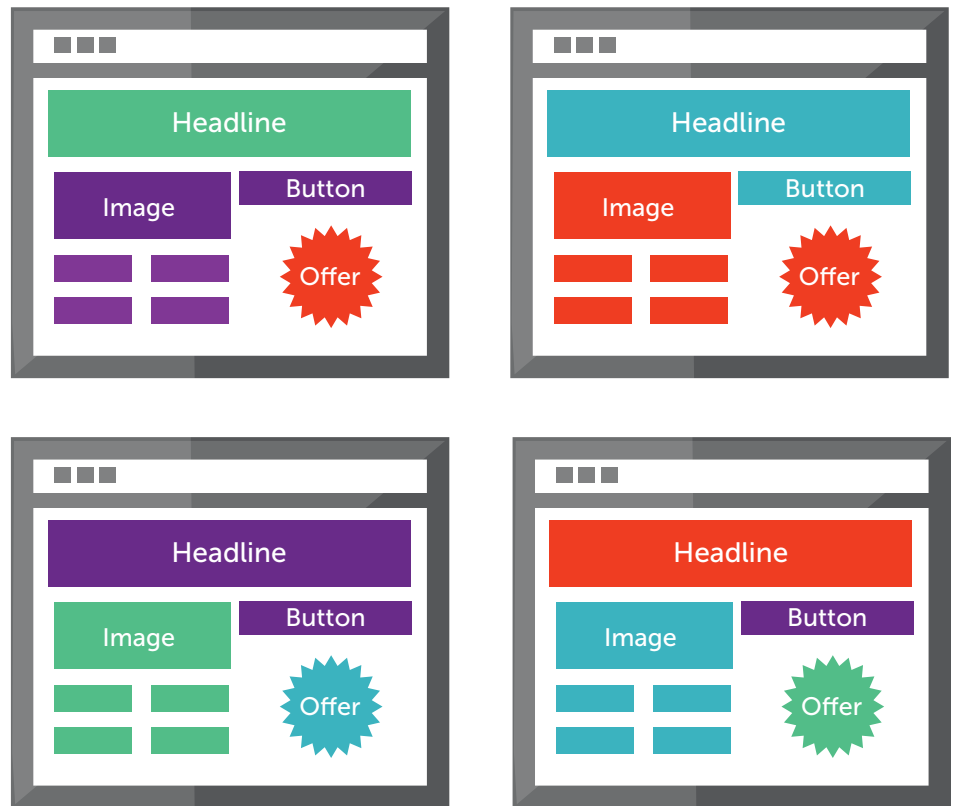


If used correctly, multivariate testing identifies the optimal layout, colors, offers, message, creative and format to improve conversions or increase revenue. It also enables the testing of individual, complex elements as opposed to an A/B test with just two page options.

However, multivariate test results can easily be ruined by poor methodology. That's why it's important to understand the rules of successful multivariate testing.

256

The illustrations show four different experiments, each representing unique combinations of levels. There actually are many more variations that could be created ($4 \times 4 \times 4 \times 4 = 256$ combinations).



Rules for successful multivariate testing

While every website optimization project has different challenges and goals, there are basic rules that apply to every multivariate test design: technical and content rules.

Technical rules

Technical rules ensure statistical significance of any multivariate test. They include three steps:

1. Choose the appropriate multivariate test type (full or fractional factorial, see sidebar on next page)
2. Determine the number of factors and levels that can be tested based on estimated conversion traffic (choose a test array or test case)
3. Stop the test when it has stabilized, not based on your earlier estimations

Running a test full factorial, if your traffic supports it, may be a good choice if you're testing content that you believe to have many interactions, or if you only want to test three factors with two levels each. Typically, you'll want to run a fractional factorial test to save time and expand the number of factors and levels you can test.

In order to find out how many factors and levels you can test, you need to have some idea of your predicted page views, conversions, and an estimate of desired lift (the upwards percentage of buyers on a targeted list). Lift matters, because a large lift will garner more conversions and your test will stabilize quicker.

To determine when a test is stabilized, look for horizontal trending of lines on graphs in your test reports. This will indicate that “winning” levels and experiments are staying winners and their level of influence or conversion rates are staying fairly constant (horizontal) over three to five days. If you don’t have graphs available, review the historical cumulative conversion rate for your experiments and see if there is a range of variance between the last few days of the test.

Factorial testing

Full factorial testing

Requires that every possible experiment combination be shown, so a five-factor test would need to display all 32 experiments. This means that if there is a sample size of 100 conversions, 3,200 conversions will be required. Since full factorial gathers additional data, it reveals all possible interactions, but there is a trade-off: More data equals more information but more data also requires a test of longer duration. The minimum data requirements for full factorial are very high since you are showing every experiment.

Fractional factorial

Displays a much smaller number of experiments. If, for instance, eight experiments were chosen, it would require about 800 conversions for a valid test. Fractional factorial multivariate testing gives you the power to make trade-offs between testing only main effects to testing for interactions based on intelligent test design. With the speed at which audiences move, marketing campaigns and seasons can change, it is important to get the most testing done in the least amount of time without sacrificing the quality of the data. Fractional factorial allows you to do just that.

Factor	Description	Level 1	Level 2
1	Image	Person	Product
2	Headline	Best MP3 Player	Listen to Music Anywhere
3	Sub Headline	Only \$49.99	Hold 1000 Songs
4	Button Text	Buy Now	I Want It
5	Main Copy	Long Copy	Short Copy

Understanding full and fractional factorial in multivariate testing

While a survey is most valid if it has 100 percent participation, it would take a long time to get “full factorial.” That’s why it is reasonable to rely on “fractional factorial” or a percentage of participants to get valid results.

Just as surveys require a certain number of participants, or sample size, for valid results, website multivariate testing works the same way.

Consider this multivariate test case:

- 5 factors = 5 parts (image, headline, etc.) and 2 levels = 2 variations.
- There are 2 levels for each factor, so you can have $2 \times 2 \times 2 \times 2 \times 2$ (2 to the 5th power) = 32 possible experiments.
- This means there are exactly 32 combinations of image shots, headlines, sub headlines, button text and main copy from the matrix.



Content rules

Content rules are closely interrelated. They ensure that the items selected for testing have purpose and that they don't needlessly expand the size of your test, reducing its efficiency.

1. Every item you test should answer an important question
2. Test variety not quantity
3. Test opposites first, then refine
4. Remember you can run more than one test

Begin designing multivariate tests by creating hypothesis regarding issues with the page, and then choose factors and design levels to address those issues. An example hypothesis is "Having a main image on the right side of the page causes users to ignore the important value proposition on the left side." To test this, choose image shot position as a factor and then have "left side image" as the baseline level and "right side image" as the second level.

This example also illustrates that, other than headlines and images, testing layout is possible with creative use of cascading style sheets (CSS) and sometimes JavaScript. As long as you can revert from one to another and it matches the other factors and levels, you are at liberty to test anything. But make sure that you are testing as few items as possible to find out what you need. For instance, before testing a collection of lifestyle images, choose one and test it against an iconic image. This will save you the time of going down a path of testing something that may not work.

Summary

Determine what you're trying to achieve, select the proper testing method to meet those goals, and then make sure to be purposeful and efficient with the content you test in front of your visitors. Follow proven testing practices rules and you'll be on your way to conquering conversion rates, bounce rates, funnel drop-offs and many other important metrics that affect your business success.

To learn more about A/B and multivariate testing with Webtrends, visit: webtrends.com/testing



webtrends®

Contact

North America

1 888 932 8736

sales@webtrends.com

Europe, Middle East, Africa

+44 (0) 1784 415 700

emea@webtrends.com

Australia, Asia

+61 (0) 3 9935 2939

australasia@webtrends.com

webtrends.com

About Webtrends Inc.

For more than 20 years, Webtrends has helped companies make sense of their customer data to drive digital marketing success. By combining innovative technology with our team of trusted and creative advisors, our solutions are designed to provide actionable insights, increase customer engagement and boost revenue.

We partner with companies at all levels of digital maturity and offer solutions in measurement and optimization. We work closely with approximately 2,000 global brands including Microsoft, KLM Royal Dutch Airlines, Kimberly-Clark, HSBC, Marks & Spencer, npower, BMW, Toyota, The Telegraph, Lastminute.com and many more.