

Should You Test That? When to Engage in SEO Split Tests

This blog was written by Tim Mehta, a former Conversion Rate Optimization Strategist with Portent, Inc.

Running A/B/n experiments (aka “Split Tests”) to improve your search engine rankings has been in the SEO toolkit for longer than many would think. Moz actually published an article back in 2015 broaching the subject, which is a great summary of how you can run these tests.

What I want to cover here is understanding the right times to run an SEO split-test, and not how you should be running them.

I run a CRO program at an agency that’s well-known for SEO. The SEO team brings me in when they are preparing to run an SEO split-test to ensure we are following best practices when it comes to experimentation. This has given me the chance to see how SEOs are currently approaching split-testing, and where we can improve upon the process.

One of my biggest observations when working on these projects has been the most pressing and often overlooked question: “Should we test that?”

Risks of running unnecessary SEO split-tests

Below you will find a few potential risks of running an SEO split-test. You might be willing to take some of these risks, while there are others you will most definitely want to avoid.

Wasted resources

With on-page split-tests (not SEO split-tests), you can be

much more agile and launch multiple tests per month without expending significant resources. Plus, the pre-test and post-test analyses are much easier to perform with the calculators and formulas readily available through our tools.

With SEO split-testing, there's a heavy amount of lifting that goes into planning a test out, actually setting it up, and then executing it.

What you're essentially doing is taking an existing template of similar pages on your site and splitting it up into two (or more) separate templates. This requires significant development resources and poses more risk, as you can't simply "turn the test off" if things aren't going well. As you probably know, once you've made a change to hurt your rankings, it's a lengthy uphill battle to get them back.

The pre-test analysis to anticipate how long you need to run the test to reach statistical significance is more complex and takes up a lot of time with SEO split-testing. It's not as simple as, "Which one gets more organic traffic?" because each variation you test has unique attributes to it. For example, if you choose to split-test the product page template of half of your products versus the other half of them, the actual

products in each variation can play a part in its performance.

Therefore, you have to create a projection of organic traffic for each variation based on the pages that exist within it, and then compare the actual data to your projections. Inherently, using your projection as your main indicator of failure or success is dangerous, because a projection is just an educated guess and not necessarily what reality reflects.

For the post-test analysis, since you're measuring organic traffic versus a hypothesized projection, you have to look at other data points to determine success. Evan Hall, Senior SEO Strategist at Portent, explains:

"Always use corroborating data. Look at relevant keyword rankings, keyword clicks, and CTR (if you trust Google Search Console). You can safely rely on GSC data if you've found it matches your Google Analytics numbers pretty well."

The time to plan a test, develop it on your live site, "end" the test (if needed), and analyze the test after the fact are all demanding tasks.

Because of this, you need to make sure you're running experiments with a strong hypothesis and enough differences in the variation versus the original that you will see a significant difference in performance from them. You also need to corroborate the data that would point to success, as the organic traffic versus your projection alone isn't reliable enough to be confident in your results.

Unable to scale the results

There are many factors that go into your search engine rankings that are out of your hands. These lead to a robust number of outside variables that can impact your test results and lead to false positives, or false negatives.

This hurts your ability to learn from the test: was it our variation's template or another outside factor that led to the results? Unfortunately, with Google and other search engines, there's never a definitive way to answer that question.

Without validation and understanding that it was the exact changes you made that led to the results, you won't be able to scale the winning concept to other channels or parts of the site. Although, if you are focused more on individual outcomes and not learnings, then this might not be as much of a risk for you.

When to run an SEO split-test

Uncertainty around keyword or query

performance

If your series of pages for a particular category have a wide variety of keywords/queries that users search for when looking for that topic, you can safely engage in a meta title or meta description SEO split-test.

From a conversion rate perspective, having a more relevant keyword in relation to a user's intent will generally lead to higher engagement. Although, as mentioned, most of your tests won't be winners.

For example, we have a client in the tire retail industry who shows up in the SERPs for all kinds of "tire" queries. This includes things like winter tires, seasonal tires, performance tires, etc. We hypothesized that including the more specific phrase "winter" tires instead of "tires" in our meta titles during the winter months would lead to a higher CTR and more organic traffic from the SERPs. While our results ended up being inconclusive, we learned that changing this meta title did not hurt organic traffic or CTR, which gives us a prime opportunity for a follow-up test.

You can also utilize this tactic to test out a higher-volume keyword in your metadata. But this approach is also never a sure thing, and is worth testing first. As highlighted in this Whiteboard Friday from Moz, they saw "up to 20-plus-percent drops in organic traffic after updating meta information in titles and so forth to target the more commonly-searched-for variant."

In other words, targeting higher-volume keywords seems like a no-brainer, but it's always worth testing first.

Proof of concept and risk mitigation for large-scale sites

This is the most common call for running an SEO split-test.

Therefore, we reached out to some experts to get their take on when this scenario turns into a prime opportunity for testing.

Jenny Halasz, President at JLH Marketing, talks about using SEO split-tests to prove out concepts or ideas that haven't gotten buy-in:

"What I have found many times is that suggesting to a client they try something on a smaller subset of pages or categories as a 'proof of concept' is extremely effective. By keeping a control and focusing on trends rather than whole numbers, I can often show a client how changing a template has a positive impact on search and/or conversions."

She goes on to reference an existing example that emphasizes an alternate testing tactic other than manipulating templates:

"I'm in the middle of a test right now with a client to see if some smart internal linking within a subset of products (using InLinks and OnCrawl's InRank) will work for them. This test is really fun to watch because the change is not really a template change, but a navigation change within a category. If it works as I expect it to, it could mean a whole redesign for this client."

Ian Laurie emphasizes the use of SEO split-testing as a risk mitigation tool. He explains:

"For me, it's about scale. If you're going to implement a change impacting tens or hundreds of thousands of pages, it pays to run a split test. Google's unpredictable, and changing that many pages can have a big up- or downside. By testing, you can manage risk and get client (external or internal) buy-in on enterprise sites."

If you're responsible for a large site that is heavily dependent on non-branded organic searches, it pays to test before releasing any changes to your templates, regardless of the size of the change. In this case, you aren't necessarily

hoping for a “winner.” Your desire should be “does not break anything.”

Evan Hall emphasizes that you can utilize split-testing as a tool for justifying smaller changes that you’re having trouble getting buy-in for:

“Budget justification is for testing changes that require a lot of developer hours or writing. Some e-commerce sites may want to put a blurb of text on every PLP, but that might require a lot of writing for something not guaranteed to work. If the test suggests that content will provide 1.5% more organic traffic, then the effort of writing all that text is justifiable.”

Making big changes to your templates

In experimentation, there’s a metric called a “Minimum Detectable Effect” (MDE). This metric represents the percentage difference in performance you expect the variation to have versus the original. The more changes and more differences between your original and your variation, the higher your MDE should be.

The graph below emphasizes that the lower your MDE (lift), the more traffic you will need to reach a statistically significant result. In turn, the higher the MDE (lift), the less sample size you will need.

For example, If you are redesigning the site architecture of your product page templates, you should consider making it noticeably different from both a visual and back-end (code structure) perspective. While user research or on-page A/B testing may have led to the new architecture or design, it's still unclear whether the proposed changes will impact rankings.

This should be the most common reason that you run an SEO split test. Given all of the subjectivity of the pre-test and post-test analysis, you want to make sure your variation yields a different enough result to be confident that the variation did in fact have a significant impact. Of course, with bigger changes, comes bigger risks.

While larger sites have the luxury of testing smaller things, they are still at the mercy of their own guesswork. For less robust sites, if you are going to run an SEO split test on a template, it needs to be different enough not only for users to behave differently but for Google to evaluate and rank your page differently as well.

Communicating experimentation for SEO split-tests

Regardless of your SEO expertise, communicating with stakeholders about experimentation requires a skill set of its own.

The expectations with testing are highly volatile. Some people expect every test to be a winner. Some expect you to give them definitive answers on what will work better. Unfortunately, these are false expectations. To avoid them, you need to establish realistic expectations early on for your manager, client, or whoever you are running a split test for.

Expectation 1: Most of your tests will fail

This understanding is a pillar of all successful experimentation programs. For people not close to the subject, it's also the hardest pill to swallow. You have to get them to accept the fact that the time and effort that goes into the first iteration of a test will most likely lead to an inconclusive or losing test.

The most valuable aspect of experimentation and split-testing is the iterative process each test undergoes. The true outcome of successful experimentation, regardless if it's SEO split-testing or other types, is the culmination of multiple tests that lead to gradual increases in major KPIs.

Expectation 2: You are working with probabilities, not sure things

This expectation applies especially to SEO split-testing, as you are utilizing a variety of metrics as indirect signals of success. This helps people understand that, even if you reach 99% significance, there are no guarantees of the results once

the winning variation is implemented.

This principle also gives you wiggle-room for pre-test and post-test analysis. That doesn't mean you can manipulate the data in your favor, but does mean you don't need to spend hours and hours coming up with an empirically data-driven projection. It also allows you to utilize your subjective expert opinion based on all the metrics you are analyzing to determine success.

Expectation 3: You need a large enough sample size

Without a large enough sample size, you shouldn't even entertain the idea of running an SEO split test unless your stakeholders are patient enough to wait several months for results.

Sam Nenner, a consultant for SearchPilot and Distilled, explains how to know if you have enough traffic for testing:

"Over the course of our experience with SEO split testing, we've generated a rule of thumb: if a site section of similar pages doesn't receive at least 1,000 organic sessions per day in total, it's going to be very hard to measure any uplift from your split test."

Therefore, if your site doesn't have the right traffic, you may want to default to low-risk implementations or competitive research to validate your ideas.

Expectation 4: The goal of experimentation is to mitigate risk with the potential of performance improvement

The key term here is "potential" performance improvement. If your test yields a winning variation, and you implement it across your site, don't expect the same results to happen as

you saw during the test. The true goal for all testing is to introduce new ideas to your site with very low risk and potential for improved metrics.

For example, if you are updating the architecture or code of a PDP template to accommodate a Google algorithm change, the goal isn't necessarily to increase organic traffic. The goal is to reduce the negative impact you may see from the algorithm change.

Let your stakeholders know that you can also utilize split-testing to improve business value or internal efficiencies. This includes things like releasing code updates that users never see, or a URL/CMS update for groups of pages or several microsites at a time.

Summary

While it's tempting to run an SEO split test, it's vital that you understand the inherent risks of it to ensure that you're getting the true value you need out of it. This will help inform you on when the scenario calls for a split test or an alternative approach. You also need to be communicating experimentation with realistic expectations from the get-go.

There are major inherent risks of engaging with SEO split-testing that you don't see with on-page tests that CRO usually runs, including wasted resources and non scalable results.

Some of the scenarios where you should feel confident in engaging with an SEO split test include where you're uncertain of keyword and query performance, proof-of-concept and risk mitigation for larger-scale websites, justification for ideas that require robust resources, and when you're considering making big changes to your templates.

And remember, one of the biggest challenges of experimentation is properly communicating it to others. Everyone has different

expectations for testing, so you need to get ahead of it and address those expectations right away.

If there are other scenarios for or risks associated with SEO split-testing that you've seen in your own work, please share in the comments below.

<http://tracking.feedpress.it/link/9375/13618519>