GUIDE TO ONLINE TESTING

AN INTRODUCTION TO TESTING FOR HIGHER CONVERSIONS

ion

ion

Example MVT Iteration Wave & Decision Tree

Iteration Wave: Testing Content Combinations

Example Combination A/B + MVT Innovation Wave & Decision Tree

Innovation Wave: Testing LX Types Against Control

Example A/B Iteration Wave & Decision Tree

Iteration Wave: Testing Segmentation Alternatives

ion

Testing Basics

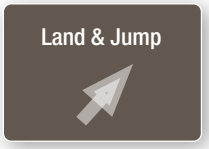
Let's start with the lay of the land. It can be challenging to understand testing without a grasp of the basics. Whether you prefer the analogy of giving the bunny the carrot or giving the gorilla the banana — the objective is the same — you're testing to find the best way to **give your user what they want**. Bunnies want carrots. Gorillas want bananas. It's simply your job to test alternative ways of giving them what they want. They tell you which alternative they prefer by taking more carrots or bananas. You're testing to provide the optimal user experience which translates — for you and your marketing — to the optimal conversion environment. Once we're all on that same proverbial page, the rest is easier…

Landing Experience (LX) Types

There are four main types of landing experiences. And since 'user experience' is often shortened to UX, we co-opt **LX for 'landing experience**' — just to keep things easy. Testing often begins with macro-level efforts aimed at identifying which type of LX works best in context. LiveBall lets non-technical people easily create any type of landing experience and test types against each other. Here are some of the most common types of landing experiences:

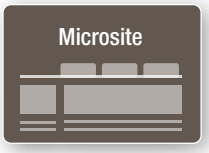
Landing Page

The good, old fashioned landing page is just that — a page. Typically, it includes the call-to-action and the method of conversion (a form). Lead-gen landing pages are typically two pages — the landing page and its post-conversion 'thank you' or confirmation page.



Land & Jump

A 'land & jump' is a truly single-page landing that matches the upstream call to action and acts as an interstitial page between the ad and the conversion. These are sometimes called warming pages as they warm the visitor prior to asking them for anything. E-commerce land & jumps often hand off visitors to either a product page or a transactional shopping cart.



Microsite

Microsites are little websites. That means they have navigation and a number of pages. Their navigation is what makes them different from conversion paths.

Conversion Path

Conversion paths are multi-page experiences (like microsites, but without navigation). They typically begin with a segmentation page that encourages visitors to group themselves by intent, sub-target, product, price, etc. Early segmentation lets a conversion path target more specifically to

visitor needs and ask for conversion later in a more specific way.

A note about home pages and deep links: Testing typically addresses dedicated landing experiences — like the ones described above. In the past, marketers have relied on website pages as entry points to their messaging. This is especially challenging in that multipurpose pages have a hard time optimizing for conversion. So technically, your home page or a deep link in your website could be your 'landing pages', but they're not dedicated landing experiences.

LX States

Control

An LX that establishes a baseline performance index is a 'control'. A parallel control runs at the same time, on the same traffic, in the same context as the LXs that are being tested against them. Sequential or serial controls are considered invalid as there are environmental differences that cloud causation.

- There can be only one control in a test wave.

Champion

An LX that wins a test by achieving superior, statistically significant conversions over the control and any other challengers is crowned champion.

- There can be only one champion in a test wave.

Challenger

An LX tested against a champion and/or a control is a challenger. If a challenger achieves a statistically significant conversion rate then it becomes the champion.

- There can be one to many challengers in a test wave.

Test Waves

A test wave or test group is a group of LXs that will receive the **same stream of traffic at the same time**. A test wave may be made up of control, champion and challenger landing experiences of varying types.

Types of Test Waves

A test wave can be characterized as either iterative or innovative. This is an important distinction as it affects both the effort necessary to produce the wave and the ability to assess causation once the wave produces a result.

Types of Testing

There are two main classes of landing experience testing. They are very different from one another and can be used alone or together (technology permitting) — you may run multivariate tests or A/B tests. LiveBall includes both A/B/n and multivariate testing technology and enables both within a point-and-click user interface. LiveBall also enables multivariate tests to be run within A/B/n tests.

A/B or A/B/n

When you test an entire LX against at least one other LX, you're A/B testing. A/B/n is just a way of noting that A/B testing isn't limited to only two alternatives.

MVT or Multivariate Testing

When you test many elements within a page — for example, versions of a headline — you are multivariate testing.

Testing Math

Some of today's testing software does a good job of insulating the test author from the statistics that are necessary to confidently predict results. That said, you still need to understand some basic concepts. Don't worry, there are no formulas in your future (assuming you're using LiveBall for your testing).

Statistical Confidence

How sure do you want to be that your test result is fact? Many testing tools declare champions with 80% confidence. That means, they are 80% confident that the champion is indeed the champion. Some tools — like LiveBall — let you specify your desired confidence level — between 80% and 99%. The higher you set your confidence, the longer it will take to declare results, but the surer you will be that your results are true.

How much time do I need to reach confidence?

The answer to this question is much more complex than it may seem. If the things you are testing are very different from one another, you may reach confidence faster if the LXs have very different conversion rates. If you have a lot of traffic, you should also reach confidence faster. But, the more similar your test subjects and the lower your traffic flow, the longer it will take for trends to turn into results. Oh, and the higher your confidence level, the longer it will take. And the more things you are testing — either A/B/n or MVT — the longer it will take.

Typically, A/B tests will reach confidence faster — because there are fewer alternatives that are very different from one another. MVT takes longer as there are typically more alternatives and the differences between them is less significant.

	Longer Test Period	Shorter Test Period
Number of Alternatives	More	Less
Difference in Results	Lesser	Greater
Traffic Volume	Less	More
Confidence Level	Higher > 90%	Lower < 90%

Understanding Testing Risk & Reward

Online optimization can be a fascinating, illuminating and rewarding undertaking. It can also be overwhelming and confusing. It's a discipline that lies at the intersection of marketing, statistics and art — three bedfellows that don't necessarily understand one another. Testing for conversion optimization should be a long-term commitment. And for that commitment to yield

success, it must meld the three disparate elements with full understanding of the inherent risks & rewards.

All Test Types are not Equal

Strategic decisions made at the onset of an optimization program have a profound impact on both the pace of the program and its results. Fundamental to this understanding is the idea that all test types do not have the same potential upside or downside.

Comparing & Contrasting A/B and MVT

A/B testing is usually more extreme than multivariate testing. Testing **experiences** against one another offers the opportunity to create wildly different things. Think in terms of testing apples and oranges. The upside of this style of testing is that it gives you the freedom to find big winners — huge lift. And the downside is that you can also find big losers. Win or lose, since so much is so different, it's unlikely that you'll be in any position to answer 'why' your apple beat your orange. Was it the shape, the color, the texture? Who knows. And for A/B testers at peace with their world — who cares? All that matters is that it won big and got big lift.

Multivariate testing is far less extreme than A/B testing. Even in cases where you are varying a number of variables, they're still limited to the context of a single **page** — in contrast to varying an entire experience with A/B. So the potential gains and losses are less extreme as well. A lot of lift can be found using MVT when the interaction effects of multiple variables come together in a powerful and perhaps unexpected way. In relatively minor MVT scenarios — where you're testing a few versions of a headline for example — you'll likely come out of your test with a pretty clear understanding of causation. In more complex cases — where many elements are varied on a page to produce a winning combination — your inference of causation may be flawed. But again, why something won or lost is less important than the reason we're testing in the first place — to generate more business from less marketing spend.

The Risk of the Endless Test

A risk to be aware of is the never-ending test. Alternatives that are too similar may not deliver a statistically significant result within a reasonable period of time. When waiting for results takes too long, a testing program can run into paralysis that is a disservice to the program investment. The purpose of testing is to get more business from less spend. Waiting months for small lifts between very similar options is not going to serve the top line purpose of the program. When you find yourself in a test wave that appears to be a statistical draw, call it as such and move on. Often, when you step back and look at your alternatives, you realize that you were timid in your approach — testing nothing of significance — playing it safe. Every test — win, lose or draw — is an opportunity to learn and improve. Waiting for statistical confidence on a test that is yielding nothing is a tremendous waste.

What kind of tester are you?

It may sound silly, but it's good to know what kind of tester you are — kind of like your testing personality type. Just as investors have varying profiles and risk tolerance, so too do online marketers in testing roles. Organizations also tend to have testing profiles — predispositions or cultural norms that affect planning, documentation and reporting. We've found that marketers exhibit characteristics that ultimately push them towards being rock stars, pragmatists or purists — and often combinations of all three.

	Rock Star	Pragmatist	Purist
View of Testing	ROI	Work	Knowledge
Pre-Test Focus	Strategy & Revenue	Creative & Production	Metrics & Attribution
Performance Granularity	Campaign	Landing Experience	Page or Element
Post-Test Focus	Next Innovation	Next Wave	Last Wave
Confidence Level	Mid (85-95%)	Low (80-85%)	High (95-99%)

Test Planning

The key to effective test planning is to have a vision of the if/then scenarios that come from your initial waves. It's counterproductive to over plan, and chaotic to under plan. At ion, we've developed a test planning framework that balances foresight with agility to consistently create high-performance programs.

Reminder: Each wave of testing shares a stream of traffic, and all test landing experiences (LX) must be run in parallel with visitors allocated at random to each LX within the wave.

Before you can design test waves, you need to divvy up your traffic streams by medium, vehicle and message. And then you need to determine if you have controls for each stream. Most of the time, tests are being plugged into existing streams of traffic with existing landing experiences — even if those existing experiences are rudimentary, like a home page or website deep link. Regardless of what your control is, you should probably test against it for wave one. You want to get to a reliable baseline and running a parallel control is the only way to get that.

Once you have an inventory of the streams of traffic and their control landing experiences, you can prioritize them by traffic volume or expense. The more volume in a traffic stream, the more potential it has to generate incremental business. And, in many cases, the larger the traffic stream, the higher the expense associated with that traffic. More traffic means there's more at stake. The quickest way to show ROI in an optimization program is to lift the performance of the most costly stream of traffic. So put them in order by traffic volume — unless there's a high-value, high-cost, low-traffic stream that's a better wave-one candidate.

For wave one, start with your highest-value traffic source and test new challenger LXs against a control that's been running on that traffic source.

Deciding What to Test

What to test in wave one depends on where you're at with your previous testing program. Some organizations have mature optimization programs while others have done little, if any, testing.

If you have a mature program, it's likely that you'll pick up where you left off with your prior efforts. ion typically recommends taking historical learnings into consideration and developing at least one innovative A/B challenger to run against your current champion (control).

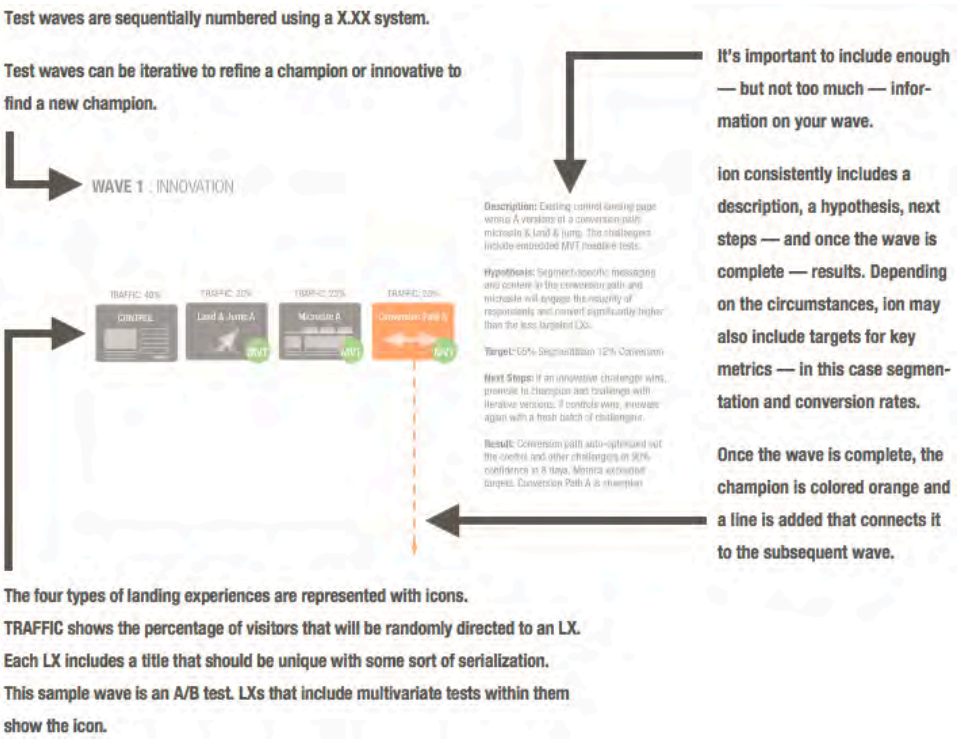
Many organizations are less developed and are moving into sophisticated testing for the first time. In those cases, it's often good to run innovative challengers against a parallel control.

Deciding How Many Alternatives to Test

The number of alternatives you choose to test — via either A/B and MVT combined — should be proportional to the amount of traffic you have. If you have a lot of traffic, then by all means, test more alternatives in a single wave. If you have limited traffic, you should be thinking in terms of one or two challengers in addition to the control or champion. Keep in mind that your overall traffic is not relevant — it's the traffic flow to the source feeding the test wave that needs to be large enough to provide an adequate sample size for each of your alternatives. What is 'adequate' depends on other factors — especially your chosen confidence level. See the section titled Testing Math for more on this.

Anatomy of a Sample Test Wave Planning Template

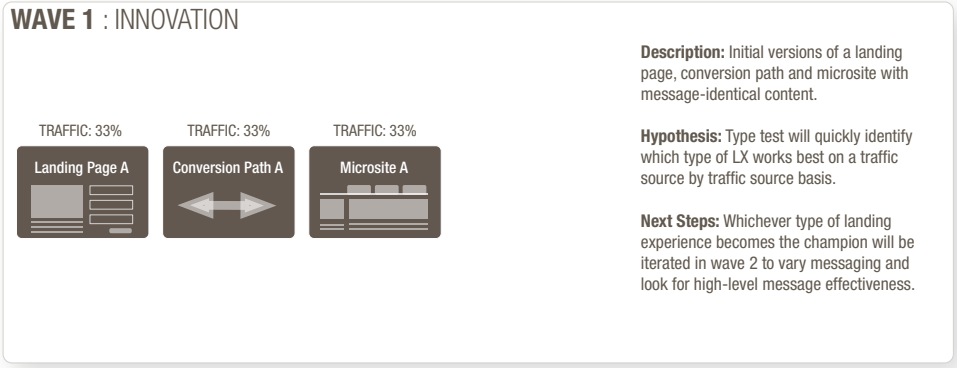
Test planning can be pretty abstract. To help with visualization and make it a bit more tangible, ion has developed a simple templating system that's explained below and used in forthcoming decision tree examples.



Test Wave Decision Trees

So here we go… We'll begin visualizing test waves and the potential outcomes and decision making that could result from those waves. The visualizations that follow are easy to reproduce using ion's template and stencil available for OmniGraffle. Email info@ioninteractive.com and request the free OmniGraffle testing templates (you'll need to get a license to OmniGraffle or OmniGraffle Pro to use the templates).

Example A/B Innovation Wave & Decision Tree



Innovation Wave: Basic LX Type Test

This is a relatively basic innovation wave — meaning it's testing wildly different alternatives. In this case, the differences are in the type of LX being tested — a landing page, versus a

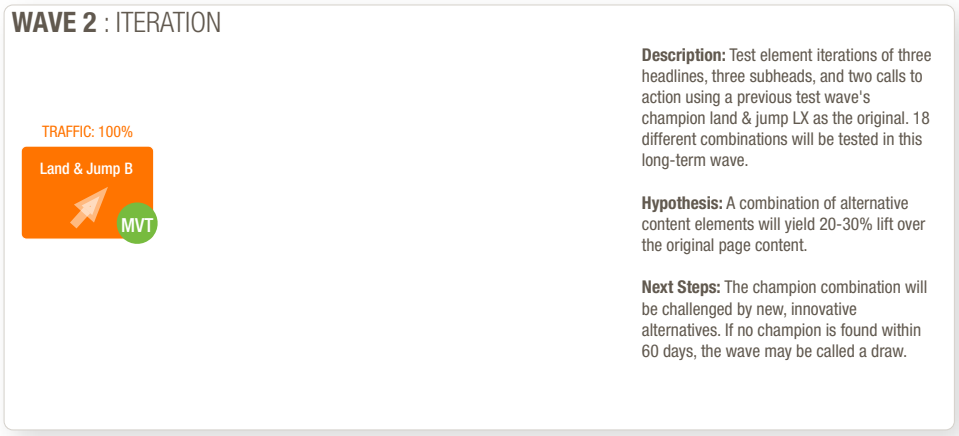
conversion path, versus a microsite. Surprisingly enough, the results of a wave like this often vary by traffic source. You may even see a microsite win in Google and a landing page win in Yahoo — on the same keyword + ad combination.

The traffic in this wave is evenly distributed amongst the LXs.

While this wave uses 'message-identical' content across the LXs, that doesn't mean it's word-for-word. Multipage LXs like conversion paths and microsites require contextually different content than a landing page. So while these appear to be apples-to-apples, they really aren't. Message identical means that the root message and offer is alike across the LXs.

If	Then
Landing Page A wins	Wave 2 : iterate landing page message or design alternatives
Conversion Path A wins	Wave 2 : iterate conversion path message or design alternatives
Microsite A wins	Wave 2 : iterate microsite message or design alternatives
Statistical draw	Wave 2 : innovate message alternatives and retest types of LXs

Example MVT Iteration Wave & Decision Tree



Iteration Wave: Testing Content Combinations

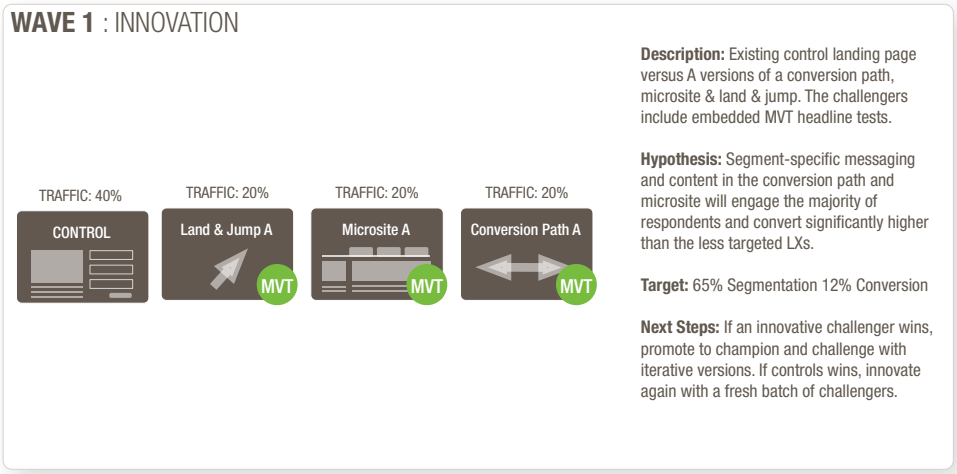
In this wave, a champion land & jump experience crowned in a previous wave is being iterated for content performance. Several versions of headlines, subheads and calls to action will be tested using full-factorial MVT to find the highest performing combination of elements.

Although this appears to be a single LX wave, it is actually a wave of 18 different combinations of content elements. Thus this will be a long-term wave requiring a good deal of traffic. In order for this wave to bear meaningful results, the alternative elements must be different enough from one another to elicit different responses from visitors. If the element variations are too similar, this wave may run for far too long without a result.

A wave like this one would likely percolate in the background while other innovation waves got more attention and effort. Assuming that the champion land & jump iterated in this wave was the result of a recent innovation, letting it take some time being refined by MVT is a good strategy. It's likely that the next wave for this source of traffic would once again focus on innovation, drawing on the learning from this wave's results.

If	Then
Content combination wins	Wave 3 : innovate with big ideas to challenge the new champion
Statistical draw	Wave 3 : reset MVT iteration with more drastic variations

Example Combination A/B + MVT Innovation Wave & Decision Tree



Innovation Wave: Testing LX Types Against Control

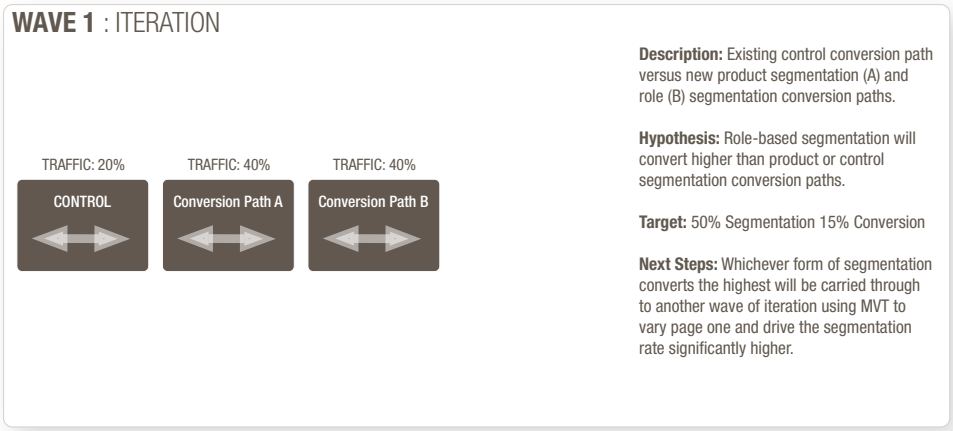
This is an innovation wave — meaning it tests significantly different LXs against one another. There are three challengers — a land & jump, a microsite and a conversion path — and a control landing page.

The traffic balance is somewhat conservative — reflecting a prudent tester. 40% of the traffic is being weighted to the control to mitigate some of the risk of the radically new challengers.

This wave is an A/B test with nested or embedded multivariate headline tests. Depending on the traffic flow and the number of MVT combinations, this wave could take quite some time to reach statistical confidence — not necessarily the best choice for a first wave looking to show testing ROI.

If	Then
Control wins and multi-page segmentation < target	Wave 2 : innovate focused on first impressions of first pages
Control wins and multi-page segmentation > target	Wave 2 : innovate focused on offer/conversion pages
Land & Jump A wins	Wave 2 : iterate land & jump challengers
Microsite A wins and segmentation > target	Wave 2 : iterate microsites with focus on offer pages
Microsite A wins and segmentation < target	Wave 2 : iterate microsites with focus on home page
Conversion Path A wins and segmentation > target	Wave 2 : iterate conversion paths with focus on offer pages
Conversion Path A wins and segmentation < target	Wave 2 : iterate conversion paths with focus on landing page
Statistical draw	Wave 2 : innovate more significant differences between LXs

Example A/B Iteration Wave & Decision Tree



Iteration Wave: Testing Segmentation Alternatives

It's often difficult for organizations to take an external view of the pain that they heal. This wave seeks to uncover how visitors frame themselves and whether role or product-centric segmentation resonates better. This wave is based on many ion has successfully used to realize significant lift with clients.

A control conversion path is tested against two challenger conversion paths. The difference between the alternatives is the axis of the segmentation options presented to visitors and the subsequent segment-specific messaging. Structurally, the LXs would be identical — hence this is an iterative and not innovative wave.

The traffic balance in this case is more aggressive as the perceived risk of the wave is lower than the previous example. And because the tester is looking for speedy results.

If	Then
An LX segments highest, but loses in conversion	Wave 2 : iterate control offer pages per conversion winner
Winner segments lower, but wins in conversion	Wave 2 : iterate control landing page per segmentation winner
Statistical draw	Wave 2 : innovate more significant differences between LXs

The Best Landing Pages Start with You. And LiveBall.

LiveBall is ion's web-based platform that empowers marketers to launch and test advanced landing pages without code or IT. If you're looking to grow your landing page program, check out these resources and find out how LiveBall can help.

Next steps for your landing pages:

If you want to stay up to-date on the latest online marketing trends and best practices, subscribe to the **ion landing page blog**.

Learn more about how you can improve your online marketing ROI by checking the free **webinars**, **white papers** and **presentations** in our library.

Join us for a **Live Demo** and see just how easy it is to use LiveBall to create conversion-focused landing pages.

Ready for more effective landing pages? **Let's talk LiveBall!**

www.ioninteractive.com

i-on interactive, inc.
200 East Palmetto Park Road, Ste. 107
Boca Raton . Florida . 33432

One Broadway . 14th Floor
Cambridge . Massachusetts . 02142

1 888 466.4332 U.S. & Canada
01 561 394.9484 International
01 561 394.9773 Facsimile