innervate

THE INNERVATE TESTING METHODOLOGY

INTRODUCTION

This document provides an overview of the ad creative testing methodology developed for and employed by Innervate.

GOALS OF THE INNERVATE TESTING METHODOLOGY

The methodology designed to support the Innervate Marketing Creative Platform enables high-velocity creative testing that drives ad creatives to perpetually improve. While creating this "upward pressure" on creative performance, the methodology also aims to reduce wasted media spend by minimizing the number of impressions required to declare – with statistical significance – when one or more ad creative(s) is inferior to the others participating in an optimization experiment.

AVAILABLE APPROACHES FOR DETERMINING CREATIVE PERFORMANCE

Within the digital marketing industry, two types of testing dominate digital optimization: Multivariate testing (MVT) and A/B/n:

MULTIVARIATE TESTING (MVT)

Multivariate testing is generally used to refine and optimize elements within an existing layout. Setting up MVT experiments is more complicated and time consuming than setting up A/B/n tests and MVT tests require more traffic (and therefore, more time) to complete a test.

A/B/n TESTING

A/B/n testing is the most popular testing method due to its simplicity and the speed with which tests yield results. Whereas MVT is useful for "tweaking" many variables within an established template, the A/B/n approach is generally adopted when marketers prefer to identify performance lifts between very different creative treatments.

THE INNERVATE MARKETING TESTING METHODOLOGY

To provide faster results with less setup time and lower traffic requirements, Innervate employs A/B/n to identify best-performing ad creatives. While Innervate uses an A/B/n approach, Innervate fully understands and capture the specific elements that drive creative performance.

For example, Innervate "knows" that Creative A differs from Creative B by its headline element, while Creative C differs from Creative B by its the background element. This hybrid approach allows users to achieve test results quickly while still providing detailed reporting on the impact of specific creative elements.

TEST CONCEPTS

The Innervate testing methodology is designed to support head-to-head testing of "creative variants" to find the best-performing creatives. One crucial prerequisite for proper creative testing is to ensure "apples-to-apples" media placements when comparing creatives. Innervate assures "apples-to-apples" creative tests across all traffic dimensionalities including, but not limited to: day part, media channel, ad format, user location, user language, user targeting parameters and trafficking duration.

Creative variants engaged in head-to-head experiments are contained within a "Creative Group." Creative Groups record and report upon:

- Test types: single element test, multiple element test, creative concept test, etc.
- Test elements: button test, call to action test, headline test, image test, etc.
- Test adjustments: color adjustment test, font size adjustment test, font type adjustment test, etc.

Creative variants proceed through multiple Test Phases where the number of Test Phases is less than or equal to n-1 the number of variants in a test group. Each phase ends when one or more creative(s) is identified as the "losing" variant.

A losing variant is determined when the Net Yield of that variant is proven to be worse than the best performing variant at a 95% confidence interval (Innervate uses a 95% confidence interval by default, but can be configured to any confidence level.) Net Yield is defined as the number of conversions attributable to a creative variant divided by the number of impressions that variant received.

Ads that participate in Test Phase 1 (with the exception of the loser) are promoted to Test Phase 2. Therefore, for the remaining phase 2 creative variants, we can compare the creative performance data cumulatively from the beginning of the test. For example, in Table 1 below, each ad is displayed as receiving 50,000 impressions in Test Phase 2. The 50,000 impression number includes 30,000 impressions from phase 1 and the additional 20,000 impressions from phase 2. By including cumulative data across Test Phases, the Innervate Marketing Creative Platform is able to reduce the time, impressions and media cost required to conclude a creative test.

Table 1 summarizes possible test results across four test phases. In this example, Creative 4 was identified – with 95% confidence — as the winning variant based upon superior Net Yield performance.

	Test Phase		Creative	Net Yield (NY)	Impressions	Conversion	Current Best	Confidence	Loser	Phase Avg. NY
TEST 1	1		1	0.0016	30,000	48	4	24%	2	0.0014
		×	2	0.0010	30,000	30	4	98%	2	
			3	0.0012	30,000	36	4	89%	2	
		✓	4	0.0017	30,000	51	4		2	
			5	0.0015	30,000	45	4	46%	2	
	2		1	0.0016	50,000	80	4	30%	3	0.0015
		×	3	0.0012	50,000	60	4	96%	3	
		✓	4	0.0017	50,000	85	4		3	
			5	0.0015	50,000	75	4	57%	3	
	3		1	0.0016	300,000	480	4	66%	5	0.0016
		✓	4	0.0017	300,000	510	4		5	
		×	5	0.0015	300,000	450	4	95%	5	
	4	×	1	0.0016	1,220,000	1952	4	95%	1	0.00165
		✓	4	0.0017	1,220,000	2074	4		1	

Table 1: Test 1, initial mean NY 0.0014, final NY 0.00165

After the completion of Test 1, additional creative variants are pitted against the Test 1 winner (Creative 4 from Table 1 above). Table 2 below shows the results of Test 2 where, over the course of five phases, creatives 4, 6, 7, 8 and 9 were eliminated leaving variant 10 as the test winner with a Net Yield of 0.0019. Note that because the Net Yields of ads 8 and 10 are so similar it took many more impressions (more than 1 million additional impressions for each variant) to declare a winner during phase five versus previous phases.

	Test Phase		Creative	Net Yield (NY)	Impressions	Conversion	Current Best	Confidence	Loser	Phase Avg. NY
TEST 2	1		4	0.0017	40,000	68	10	50%	9	0.00163
			6	0.0016	40,000	64	10	69%	9	
			7	0.0015	40,000	60	10	83%	9	
			8	0.0018	40,000	72	10	26%	9	
		×	9	0.0013	40,000	52	10	97%	9	
		✓	10	0.0019	40,000	76	10		9	
	2		4	0.0017	80,000	136	10	65%	7	0.0017
			6	0.0016	80,000	128	10	85%	7	
		×	7	0.0015	80,000	120	10	95%	7	
			8	0.0018	80,000	144	10	36%	7	
		✓	10	0.0019	80,000	152	10		7	
	3		4	0.0017	150,000	255	10	80%	6	0.00175
		×	6	0.0016	150,000	240	10	95%	6	
			8	0.0018	150,000	270	10	48%	6	
		✓	10	0.0019	150,000	285	10		6	
	4	×	4	0.0017	340,000	578	10	95%	4	0.0018
			8	0.0018	340,000	612	10	66%	4	
		✓	10	0.0019	340,000	646	10		4	
	5	×	8	0.0018	1,360,000	2448	10	95%	8	0.00185
		✓	10	0.0019	1,360,000	2584	10		8	

Table 2: Test 2, initial mean NY 0.00163, final NY 0.00185

Table 3 below shows the results of a five-phase test where creative 10 from Test 2 is entered into an optimization experiment competing against creatives 11, 12, 13, 14 and 15. In this test, creative 10 was eliminated in the fifth phase and creative 11 was declared the winner with 95% confidence.

	Test Phase		Creative	Net Yield (NY)	Impressions	Conversion	Current Best	Confidence	Loser	Phase Avg. NY
TEST 3	1		10	0.0019	30,000	57	11	42%	13	0.00173
		✓	11	0.0021	30,000	63	11		13	
			12	0.0017	30,000	51	11	74%	13	
		×	13	0.0014	30,000	42	11	96%	13	
			14	0.0018	30,000	54	11	60%	13	
			15	0.0015	30,000	45	11	92%	13	
	2		10	0.0019	40,000	76	11	47%	15	0.0018
		✓	11	0.0021	40,000	84	11		15	
			12	0.0017	40,000	68	11	81%	15	
			14	0.0018	40,000	72	11	66%	15	
		×	15	0.0015	40,000	60	11	95%	15	
	3		10	0.0019	90,000	171	11	66%	12	0.00188
		✓	11	0.0021	90,000	189	11		12	
		×	12	0.0017	90,000	153	11	95%	12	
			14	0.0018	90,000	162	11	85%	12	
	4		10	0.0019	160,000	304	11	79%	14	0.00193
		✓	11	0.0021	160,000	336	11		14	
		×	14	0.0018	160,000	288	11	95%	14	
	5	×	10	0.0019	370,000	703	11	95%	10	0.002
		✓	11	0.0021	370,000	777	11		10	

Table 3: Test 3, initial mean NY 0.00173, final NY 0.002

Figure 1 below shows the improvements to average creative Net Yield over the three tests described above (each test is a different color) with each test comprising phases (each node below shows the Net Yield improvements from each Test Phase). The x-axis in Figure 1 represents Test Phases.

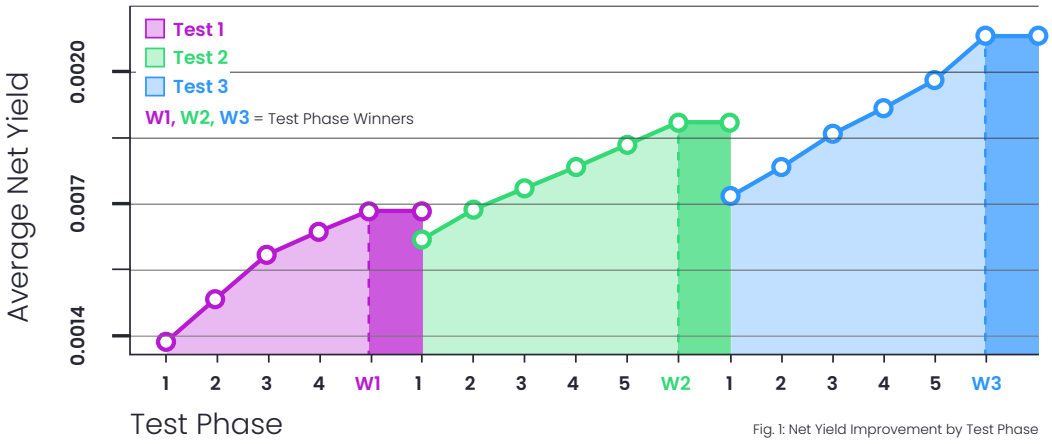

Fig. 1: Net Yield Improvement by Test Phase

The darkly shaded area in Fig. 1 above represents the period when the best performing creative from the previous Creative Group runs unopposed. In this example, as new tests are deployed, average Net Yield drops initially until the sub-optimal creatives from the next Creative Group are eliminated in subsequent phases.

In figure 2 below, the x-axis represents not Test Phases, but the average number of impressions per creative required to complete a phase (once a losing creative was identified with 95% confidence). As discussed during the analysis of Table 2, the last phase of the second test required many more impressions to determine a losing creative compared to any previous test phase. In this case, Innervate could've been configured to execute logic that ends a test after a specific amount of time or reduces the threshold of statistical significance required to determine a losing creative. Or, an Innervate user could set notifications alerting her to manually end a test and promote one or both of the top creatives to a new test. The goal of both the automated and manual approaches is to limit the opportunity cost (in terms of time and Net Yield) of delaying another round of promising creative tests.

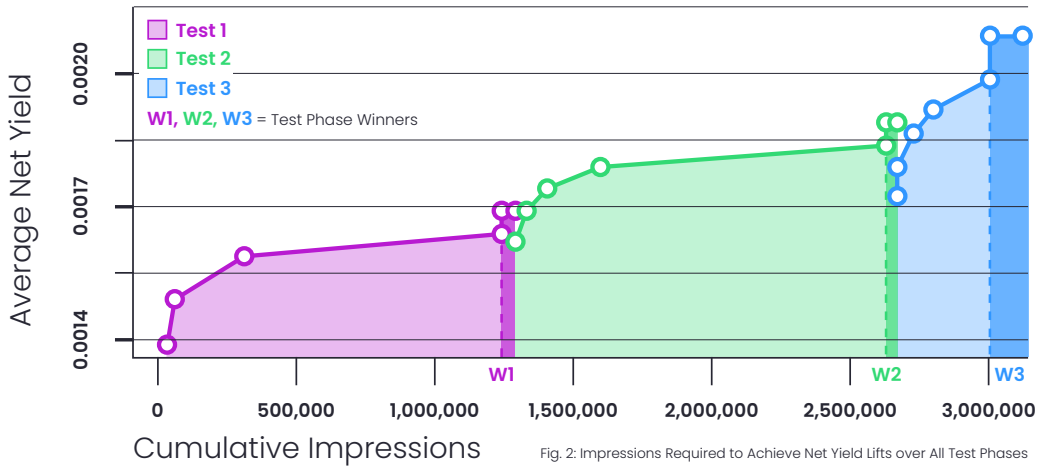

Fig. 2: Impressions Required to Achieve Net Yield Lifts over All Test Phases

The darkly shaded area in Fig. 2 above represents the period when the best performing creative from the previous Creative Group runs unopposed. In this example, as new tests are deployed, average Net Yield drops initially until the sub-optimal creatives from the next Creative Group are eliminated in subsequent phases.

CONTRASTING THE TYPICAL A/B/N APPROACH WITH THE INNERVATE MARKETING CREATIVE PLATFORM A/B/N APPROACH

Over the course of a test, A/B/n optimization tools on the market today typically run variants in head-to-head competition until a statistically-significant winner emerges. With this approach, even those variants known to be sub-optimal are included in the test rotation for the entirety of the test which results in flat average Net Yields over the test run, followed by brief periods where the winning creative runs unchallenged (the darkly shaded area of Fig. 3 below.) In an advertising context (where media costs are present), running sub-optimal ads during the test run results in wasted media spend. As figure 3 below illustrates, Net Yield improvements — and thus, improved return on ad spend (ROAS) — are only realized after the test finishes.

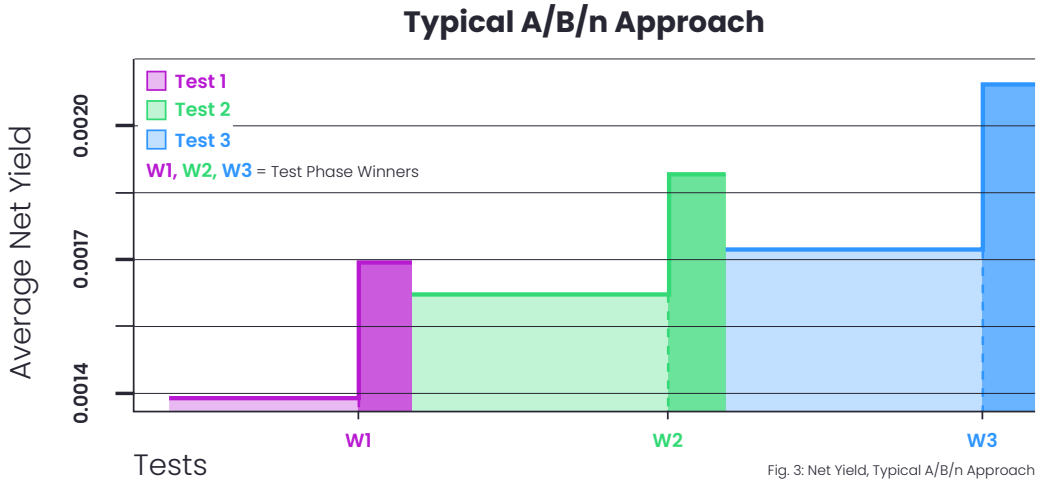

Fig. 3: Net Yield, Typical A/B/n Approach

In contrast, because the Innervate approach to A/B/n testing pauses poorly-performing creative in phases *during the test run,* Net Yield improvements are realized incrementally at the end of each Phase, not just at the end of a test (see figure 4 below). This approach allows Innervate users to capture both the incremental Net Yield improvements while the experiment runs and as well as during the period when the winning creative runs un-opposed (see the darkly shaded area of Fig. 4 below).

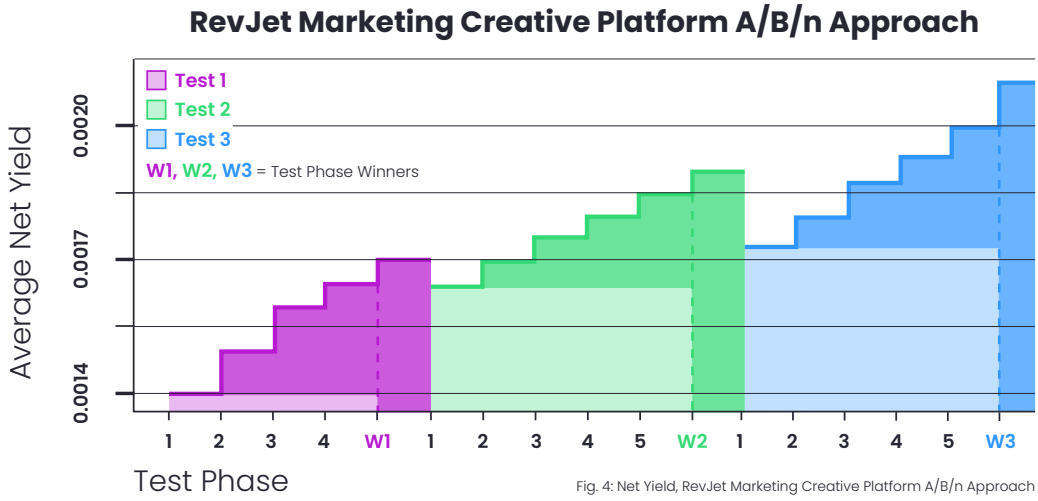

Fig. 4: Net Yield, RevJet Marketing Creative Platform A/B/n Approach

CONFIDENCE SCORING

Users employing optimization methodologies always face a dilemma with regard to confidence scoring: "How much data is enough?" Declaring winners too soon potentially leads to false positives, while waiting for more data by definition results in longer feedback loops and sub-optimal media spending on underperforming creatives. The Innervate methodology is not designed to achieve statistical certainty, nor to accurately measure the exact amount by which a winning ad outperforms a losing creative. In its efforts to boost results, the Innervate testing methodology is designed to achieve two goals:

1. To determine that an underperforming creative is indeed a "loser" and not a product of chance.
2. To limit financial losses associated with investing in underperforming creatives as quickly as possible.

Although the Innervate test confidence interval is entirely configurable by the user, by default the system requires a 95% confidence interval to remove "losing creatives" from a Test Phase. Because there exists no "correct" level of statistical significance, choosing a confidence interval requires making a trade-off between limiting false positives and getting fast results. In our experience, 95% is the most appropriate confidence level that balances the desire to reduce false positives and with the goal of quickly receiving test results.

Innervate employs a two-tailed z-test to distinguish winning creatives from losing creatives. The z-score employed by Innervate is calculated as:

$$Z = \frac{P - P_c}{\sqrt{P_P\left(1 - P_P\right)\left(\frac{1}{N} + \frac{1}{N_c}\right)}}$$

WHERE

P	Proportion (test)	Net Yield of test ad
P_c	Proportion (control)	Net Yield of control ad
P_P	Proportion (pooled)	Combined Net Yield of both test and control ads
N	Sample size (test)	Impressions of test ad
N_c	Sample size (control)	Impressions of control ad

SUMMARY

This overview covered the following key aspects of the Innervate testing methodology:

- The goals of the Innervate testing methodology are: 1) to continually improve the performance of digital ad creative via high-velocity creative testing, and 2) to minimize the media budget (and number of impressions) required to test and improve creative performance.
- The Innervate testing methodology employs an A/B/n approach to optimization while fully retaining the ability to report on the impacts of specific ad elements (e.g., headline tests, color tests, font tests, call to action tests, etc.)
- The Innervate methodology ensures "apples-to-apples" comparisons of creatives across traffic dimensionalities (e.g., day part, channel, format, trafficking duration, etc.)
- Creative variants compete within Creative Groups. Over Test Phases, sub-optimal creatives are eliminated when they are found to perform with a significantly lower Net Yield than the winning creative.
- Rather than running an entire test with sub-optimal creatives, Innervate eliminates poorly-performing creatives during Test Phases, which reduces wasted media spend.
- The Innervate methodology employs a two-tailed z-test to distinguish winning and "losing" creatives.
- Although completely configurable, by default Innervate uses a 95% confidence level to determine the differences between creative performances.

Learn more today and contact us at
info@Innervate.com or visit www.Innervate.com