DATAPOP

A COMPREHENSIVE GUIDE TO THE ART & SCIENCE OF PPC AD OPTIMIZATION

Table
of contents

Why Do
Ad Optimization?

> "Ad copy is the only part of your AdWords account a consumer will ever see. When a searcher sees your ad, the ad needs to stand up, wave its arms, jump up and down, and shout "I have the answer to your question!"[1]
>
> **Brad Geddes**

For many search practitioners, optimizing ad copy typically sinks to the bottom of the priority list, if it makes the list at all. Most often, this is because ad optimization is challenging to do well, not because it isn't believed to be critical.

If you spend most of your time managing keywords and bids, it is easy to forget that ad copy is the only part of your account that is visible to consumers. Because it acts as the storefront of your account, you should focus your efforts on ensuring that it is current, relevant, and click-worthy at all times.

The value of ad optimization, however, goes beyond appearance. It is a key driver of performance, and, just like keyword selection, bid management, account structure, and landing page optimization, it has a significant impact on the overall performance of the account.
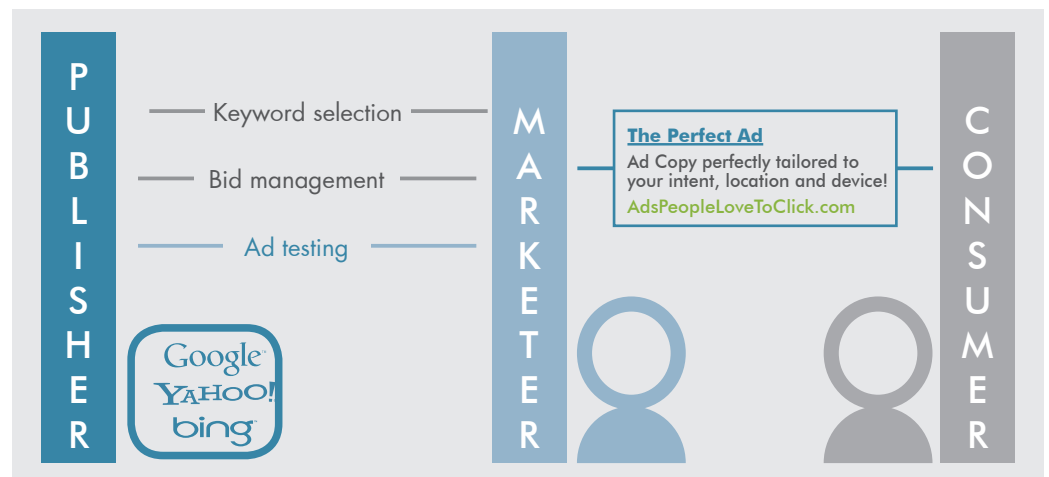
> "Ad copy is probably the most important and most challenging area, one that often takes a backseat in campaign managers' arsenals to the things users cannot see, like keywords and ad group structure."[2]
>
> **Andrew Goodman**

Recognizing the importance of ad optimization and the value it can drive for your business is the first step. However, there are still hurdles, such as the lack of sophisticated testing tools or the difficulty of scaling tests across your account. Additionally, it tends to be challenging to include ad optimization as a prioritized, structured and scheduled task in the overall account management process.

Despite these challenges, the potential impact of consistent ad optimization is an extremely valuable asset to your account, and it is crucial to take the time and effort to approach it in a structured and systematic way. High quality ad optimization is the product of both creative and mathematical processes. The art represents the actual creation of the ad copy. The science focuses on using a clean and statistically relevant testing methodology, as well as analyzing the data in a way that ensures it can be acted on appropriately.

This whitepaper will examine both the art and the science of ad optimization. It should serve as a detailed guide to the development of ads that drive significant, sustainable performance gains for your accounts.

The Art:
Building Great Creative

> *Good ads convey the information that the searcher requested, while bad ads ignore their request entirely.*

Why Most Ads Miss The Mark

More bluntly stated, search ads generally suck. You know it and so do increasingly savvy searchers who have been bombarded with misplaced and irrelevant messages for years.

Ball of Soccer at Amazon
Buy **ball** of soccer at **Amazon**!
Qualified orders over $25 ship free
www.amazon.com

Great Deals on eBay
Huge selection of great items
Free Shipping available. Buy Now!
www.ebay.com

Floral Dress
Find Floral Dress Online.
Free Shipping $50 on 100,000 items
www.shopretail.com/FreeShipping

Fascists
Great deals on **Fascists**
shop today on Official eBay site.
www.ebay.com

Flights to Hell
Find Low Prices & Great Deals On Airfares. Browse Promotions Online!
www.SmarterTravel.com

A glance at the ads above, though amusing, reveals ad copy faux pas patterns commonly practiced by even the most sophisticated advertisers.

The most recurring are:

1. Partial or total irrelevance of ad to intent
2. Robotic tone, usually amplified by misuse of DKI
3. Lack of value proposition and key differentiators

> *A clear and specific call-to-action can separate a qualified from an unqualified click, ultimately helping to deliver a strong prospect to your site or saving you the cost of a click.*

Key Elements of a Great Ad

It's quite easy to spot bad ads, especially if they are similar to the ones above. But even the worst ads can be remedied by understanding the various components that go into making a good one. The following are the top three elements that make an ad great, developed by Bryan Eisenberg and often referred to as the Conversion Trinity[3].

1. Relevance

When it comes to online advertising, relevance is more important than bidding. A relevant ad is one that reflects the search query that triggered it, while also taking signals like device, location, and context into account. Searchers naturally click on ads that closely match the language that they used in their queries as well as those that match their state of mind. If they have searched for "software," they would be more likely to click on an ad that contains the term "software" than one that highlights a "tool" or "platform." A relevant ad also provides information that helps a user in their decision process. Good ads convey the information that the searcher requested, while bad ads ignore their request entirely.

2. Value

The value communicated in an ad should be reflective of the fundamental strengths and differentiators of your business. For instance, if your market position is providing the best customer service or the lowest prices, it is critical to communicate that in your ads. With congested SERPs, close proximity to competitors and often identical looking ads, displaying your Unique Value Proposition (UVP) is critical in making your ads stand out. Expressing a proper value proposition goes beyond helping you not just blend in with competitors. It is also reflective of the experience you deliver on your site or in your stores, and ultimately communicates to searchers why you are the one that will deliver an answer to their question once they click.

3. Call-to-Action

The importance of Calls-to-Action lie in the fact that they act as predispositions for what searchers are expected to do on the landing page. Naturally, conversion rates dramatically increase when a searcher knows what to expect when they click on the ad. More importantly, though, a clear and specific call-to-action can separate a qualified from an unqualified click, ultimately helping to deliver a strong prospect to your site or saving you the cost of a click.

A searcher looking for a free demo is unlikely to click your ad if your call to action is "Download Whitepaper."

Calls-to-Action are also a key area where you can display your true understanding of the searcher's intent. CTAs can separate the searchers (Learn More) from the comparison shoppers (Compare Prices) from the buyers (Shop Now), and putting the right CTA in place will demonstrate to a searcher that you understand them and want to respond to their specific request.

Relevance

CONVERSION
TRINITY

Call to Action   Value
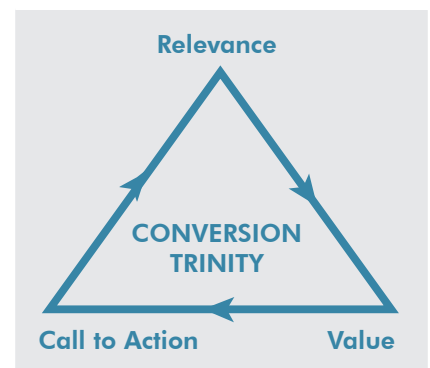
> *When it comes to dynamic data, bid tools optimize bids in near real-time, display ads render in real-time, and landing page tests come together in real-time. There's no reason PPC ads shouldn't reflect the most current state of your business as well.*

Moving Towards The Ideal

From headlines to value propositions to URLs and CTAs, there is certainly no shortage of tests that you could run.
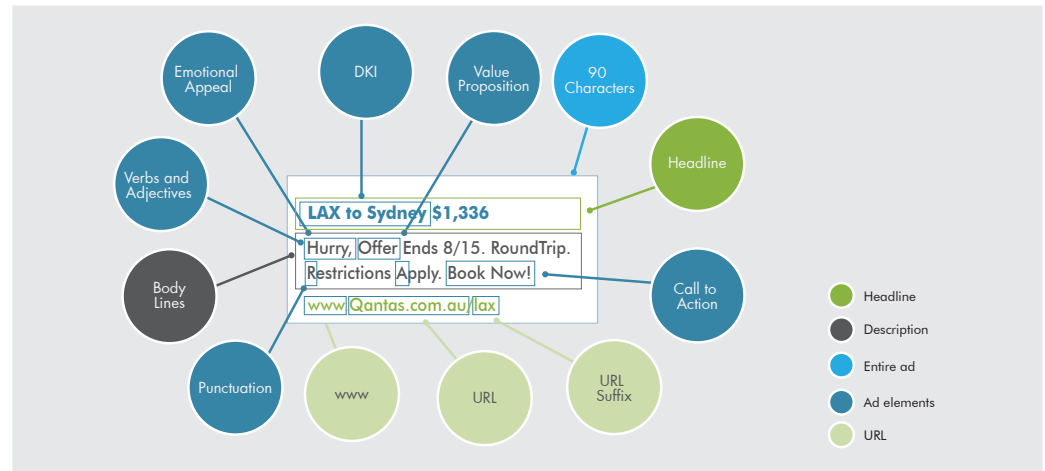
The following are some critical elements you should be testing at some point.

Headlines

Headlines are typically the first element a searcher sees in an ad. In those critical 25 characters, you must convince the searcher that you are interesting, and that they should continue to read the rest of the ad to learn more. There are multiple techniques available to you to leverage in headlines and these could ultimately be your most impactful tests.

Benefits vs. Features

The dilemma of focusing on benefits or features is a challenging one and most often, the answer can only be revealed through testing. Whether the ad copy should be more logical and factual or emotional usually depends on where the searcher is in the conversion funnel.

The top of the funnel is more awareness-driven, and therefore, informational ad copy tends to work best. The middle of the funnel, which falls around interest, usually attracts searchers that respond best to messages encouraging evaluation and comparison. The bottom of the funnel is more action focused and requires ads that highlight a reason to buy from you today with a clear CTA.

That's not to say, however, that there is an easy answer to what messages your ads should emphasize. It is crucial to mix benefits and features while testing to determine if one type of ad copy resonates better with your target searchers.

Testing feature only ads, benefit only ads and feature-benefit mixed ads against each other will help to identify which combination will resonate best with searchers throughout the funnel.

The following is a sample benefits and feature ad:

Feature:	Vibrating Watch Alarm
Benefit:	Get pill reminders even if you have trouble hearing
Feature + Benefit:	Vibrating pill reminder. Works even when Mom's TV is blaring away
	Vibrating pill reminder works even if Dad's hearing aid is off

Testing feature only ads, benefit only ads and feature-benefit mixed ads against each other will help to identify which combination will resonate best with searchers throughout the funnel.

Unique Value Propositions (UVP)
You may have multiple value propositions that tend to resonate with your customers. What counts as a value proposition? Everything from low prices, to great customer service, to free whitepapers, to wide selection, to specific product or service benefits or features. In different scenarios, different UVPs will resonate with searchers and drive the best results. It is key to figure out the different optimal combination of UVPs that resonate with searchers in different stages of your funnel.

Call-to-Action (CTA)
There are countless CTAs that you can use, but the goal of testing is to figure out which one you should use, based on the intent of your consumer. A CTA that invites people to 'Learn More' about a specific product should likely be tested during what appears to be a consideration phase, while the harder sell of 'Buy One Today!' should likely be reserved for low funnel shoppers. Figuring out which type of CTA is just a first step. There are also variations within each type that you can test, such as 'Buy One Today' or 'Shop Online'.

Dynamic Data
Bid tools optimize bids in near real-time, display ads render in real-time, landing page tests come together in real-time...so there's no reason PPC ads shouldn't reflect the most current state of your business as well. Google has started to give advertisers tools to test live pricing, inventory levels, expiration dates, and other dynamic attributes in real-time. You should start to take advantage of these options as timely and accurate data can make an ad uniquely engaging on a SERP and drive higher conversion rates on the landing page.

Promotions
Many advertisers have multiple promotions running across their business that could be incorporated into their ad messages. Many of these are not leveraged appropriately due to the limited resources applied to ad testing. The inclusion of promotional information often can be the difference between someone buying from you or your competitor.

Summary of ad elements to test:

Headline	Includes Price	Shoes From $99
	Includes Brand	Brand® Shoes
	Includes CTA	Shop Now, 100s of Shoes

Benefits vs Features	Benefit	Run faster
	Feature	Lightweight shoes
	Benefit+Feature	Run faster with light shoes

Unique Value Propositions	Selling points	Free shipping, low prices
	Product/Service Features	Colors, styles, wifi
	Experience	Wide selection, luxury service

Call-to-Action	Research	Learn more, see selection
	Evalution	Compare prices, choose plan
	Purchase	Shop today, get tickets

Dynamic Data	Price & Urgency	Starts at $125, Ends 12/31
	Inventory	Only 12 left in stock
	Geo Information	Flights from LA to NYC

Promotions	% off	10% off, end 1/31
	Seasons	Black Friday deals
	Discount	$50 off $100 purchase

- What to test
- Sample strategies
- Examples

When deciding what to test, it is important to keep in mind that testing higher impact and more visible elements usually yields high-impact results. For example, testing a headline would typically have a higher impact than testing the removal or inclusion of "www" in the URL. However, there is still potential for more subtle elements to impact performance, and drive insights for future growth.

'*Proper ad testing begins by first identifying the right areas in the account to test, which ad strategies and elements you'd like to test, and finally, how to conduct those tests.*

As you get eager to test, you might find yourself running multiple ads in many ad groups and seeing how the results play out. Whereas you may occasionally get lucky with this method, proper ad testing begins by first identifying the right areas in the account to test, which ad strategies and elements you'd like to test, and finally, how to conduct those tests.

Where to Test

There are many areas of an account that could benefit from testing, and that can offer insightful learnings on what messages drive performance. Many marketers focus high volume ad groups, but it is also important to note that even less prominent ad groups can still contribute to account performance, especially when taken in aggregate.

A good starting point for your optimization efforts is identifying the best and worst performing ads as measured by your predefined success metric. This identification will enable both action and learning. After looking at your ad performance and identifying patterns or performance impacts, you can decide which areas of the account to focus on and in which order.

You may uncover a handful of underperforming ads. You can pause, delete (we recommend pausing to save history), or optimize them. Most importantly, focus on understanding why they underperformed. High performing ads should not be ignored either, since there is always room for optimization and improvement.

If you are uncertain about where to start testing, the following four categories may provide a good starting point, as they tend to be reasonably easy to identify within the account:

1. **High-Impact Ad Groups**, as they represent the low hanging fruit and where you could see the biggest bang for your optimization efforts.

2. **Low Quality Score Ad Groups**, as they impact ad rank and the overall quality of your account.

3. **Poor CPA Ad Groups**, as these could be dragging down overall account performance.

4. **Ad Groups Representing Specific Intent**, if you have a strong theory about what types of searchers could be most impacted by your test.

As a best practice, you should identify a certain percentage of the traffic in your account and plan to constantly be optimizing across that percentage at any given time. Another way of going about that is identifying a certain number of tests to run on a monthly or quarterly basis. This will depend on your comfort level, taking into account considerations such as account size and available resources.

> *The purpose of a test is not to prove a hypothesis right, but rather to use it as a framework for executing the test and a reference point to understand the results.*

How to Test

130 characters and your desktop tool are not entirely sufficient to create a successful ad test. Most tests do not achieve desired long term results because one or more mistakes are made in the design and/or execution of the test:

The 7 Deadly Sins of Ad Optimization

1. Lack of or Poorly Developed Hypotheses
Every test must begin with a clear, specific, and informed hypothesis. Most times there isn't a hypothesis because it is easy enough to write an ad and set it live to see how it does, and a forward thinking hypothesis is overlooked. Other times a hypothesis is not used because it seems difficult or pointless to prove a hypothesis right. However, the purpose of a test is not to prove a hypothesis right, but rather to use it as a framework for executing the test and a reference point to understand the results.

2. Speaking to Multiple Intents in Test Groups
Many tests are conducted on ad groups that represent multiple user intents, making it incredibly difficult to interpret the results in a way that could drive long term value. There are two ways this happens. First, an advertiser will include keywords like "red shoes" and "red sneakers," in the same group, and as such the winning ad tends to be the least common denominator, not the best ad for each of those intents. Second, an advertiser will run their test on a broad match term with no match modifications or negative keywords. The winning ad is typically non-specific and the results of this test are largely unrepeatable.

3. Overlooking Variations in Traffic & Devices
When it comes to traffic considerations, many seem to be a lot more concerned with having enough traffic and less concerned with traffic fluctuations and variations. Factors such as seasonality leading to traffic dips and spikes can significantly affect the results of a test and the insights gained from them. Device considerations are also crucial in ad testing. Because searchers respond differently to ads when on desktops than when on mobile, you must be aware that both the ad copy and test results cannot be indiscriminately duplicated and applied across devices.

Consumer Behavior Varies Across Device

SCREEN	PC	TABLET	PHONE
TIME	Weekday heavy	Down time browsing and shopping	Always with consumers
INSIGHTS	Task oriented	High average order sizes	Optimal for local and personalized ads

4. Excessive Variation Between Test and Control Ads

One of the caveats of creating a test ad that is excessively different from the control is that when analyzing the results, whether the test ad was a winner or loser, it is nearly impossible to determine which factor was responsible for the performance swings. This is why it is crucial to be as specific as possible when creating a hypothesis and structuring a test around it. The experimental ads should be created in a way that allows you to test the variables you intend to test, and the results can then be attributed to predefined changes.

5. Declaring Tests Over Too Soon

Testing can either be exciting or terrifying, depending on what the initial results begin to look like. As the test begins to accumulate data, many make premature assumptions about which ad is a winner and which is a loser. An example is not giving an ad as long as a sales cycle typically is before making a call on the test, which doesn't give enough time for an accurate conversion rate to be revealed. Test results take time to normalize as they accumulate more data and gain increased accuracy as a result. Declaring a test over too soon before results normalize and before it has reached statistical significance means that decisions were made based on erroneous, incomplete data.

### 6. Applying Results Blindly Across Campaigns

The success of a specific element in one ad group, a 'Free download' CTA, for instance, is by no means an indication that it would yield the same results if applied across all other ad groups and campaigns, particularly those representing different intents. The only way to determine if leveraging results across other ad groups and campaigns would yield the same success is by testing before applying.

7. Too Much Focus on Brand Terms

Every marketer must own their brand in search, but it is a problem if the majority of ad testing is done on brand terms and then the results are assumed to be applicable to non-brand terms. A consumer who seeks out your brand specifically should be messaged to much differently that someone who is not as familiar with your brand. Trying to combine test results from brand and non-brand searches sets you up to miss the mark with a large portion of your ad creative.
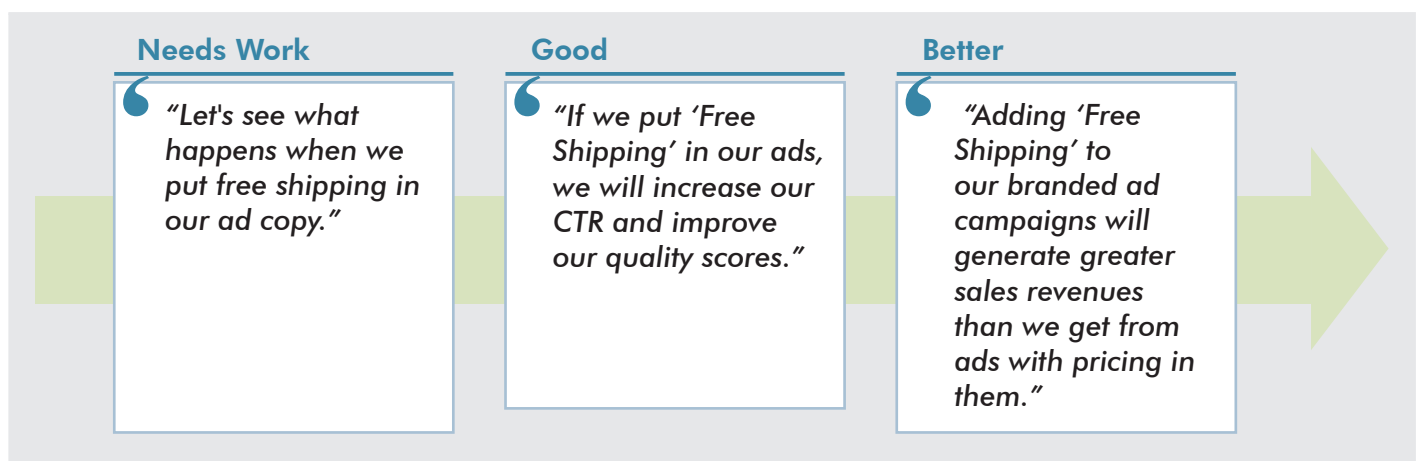
Creating a Solid Hypothesis

To create a good hypothesis, start with a question you are curious about. For instance, *"why aren't people clicking my ad?"*. The answer to this question could be similar to *"maybe because I don't have a free trial offer."* Your hypothesis, then, is the outcome you expect if you implemented your answer through a specific change. More precisely, your hypothesis in this case could be *"Adding a free-trial offer in my CTA will increase my ad's CTR"*

Question

" Why aren't people clicking my ad?

Answer

" Maybe because I don't have a free trial offer

Hypothesis

" Adding a free trial offer in my CTA will increase my ad's CTR
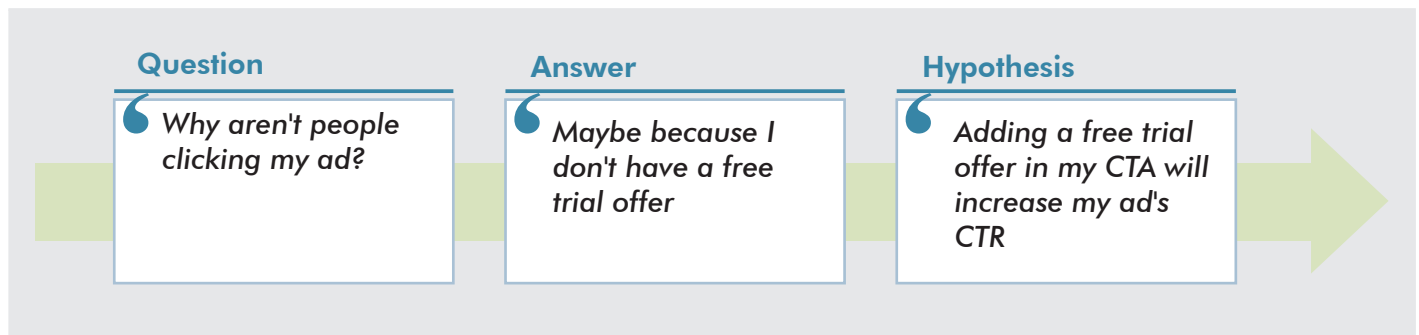
Other themes or ad components that could be used to formulate a hypothesis include:

- **Promotional Language -** e.g. does *'free shipping for $50 spent'* or *'10% off all shoes'* drive a higher conversion rate for shoe shoppers?

- **Branding Language** - e.g. does *'Official BMW Site'* drive a higher CTR than *'The Official Site of BMW'*?

- **Dynamic Data** - e.g. if I use the actual ticket price in my ad a week before the event, will ticket sales go up?

- **Headlines, Descriptions, Display URLs** - e.g. will CTR increase if I include the price in the headline versus the description?

The more specific the hypothesis is, the easier it will be to create a clear framework for testing. Here is an example of how you could quickly improve a hypothesis:

Needs Work

" *"Let's see what happens when we put free shipping in our ad copy."*

Good

" *"If we put 'Free Shipping' in our ads, we will increase our CTR and improve our quality scores."*

Better

" *"Adding 'Free Shipping' to our branded ad campaigns will generate greater sales revenues than we get from ads with pricing in them."*

Developing a good hypothesis is not random and should be informed by some degree of knowledge, irrespective of any certainty around what the outcome may be. Best practices provide a great starting point for developing hypothesis, however they don't always yield positive results. You should be especially wary of absolutes such "free trials always improve CTR!" It is important to be skeptical and to test when determining whether absolutes are applicable to your unique set of products or services presented in the context of your ads.

> *In most ad tests, more than one metric will change. While it is important to keep a pulse on all main metrics, it is crucial to identify and focus on the one that is most aligned with your overall business goals.*

Defining Success Metrics

When choosing a success metric to optimize for, it's important to keep original business goals in mind and determine which metrics align best. Most often, more than one metric changes during a testing experiment. So while it is important to keep a pulse on all main metrics, it is crucial to identify and focus on the one that is most aligned with your overall business goals.

The following is a list of the most common ad optimization metrics along with some thoughts on why they might or might not be appropriate for your business.

Click Through Rate (CTR) is the most common success metric used in ad optimization, as it is the most direct link between the consumer target selected and the quality of the ad constructed for that target. Additionally, the sample sizes for CTR are much higher than for conversion metrics which leads to faster test results. Keep in mind, however, that just because an ad has high CTR, does not mean it will necessarily convert well. Many advertisers make unreasonable promises to consumers in ad copy and these types of ads typically come at the expense of conversion rates, even though they deliver the highest CTR.

Conversion Rate (CR) is often used by advertisers focused on actions taken on their site including sales, lead generation, downloads, video plays and sign-ups. The potential downside of strictly using Conversion Rate as a success metric is that it doesn't take into account CTRs and click volumes. Despite a rise in CR, a drop in CTR or clicks could impact campaign performance negatively by leading to lower conversion volume or higher CPCs. Additionally, ads are typically not the only determinant of CR so be careful before choosing this as the target metric for your tests.

Cost Per Acquisition (CPA) similar to conversion rate in that it doesn't take into account click volumes or CTR, but since cost is a variable, it does take into account the CPC benefit of a higher quality score. This is typically a more attractive metric to use than CR when doing ad optimization.

Return on Ad Spend or Investment (ROAS/ROI) is frequently used as a success metric by performance marketers in retail. It doesn't take into account click volumes but should give a better sense of CTR impact as higher QS/lower CPC should drive a higher ROI.

Conversions Per Impression (CPI) takes both CR and CTR directly into account. Very often, increasing the CTR of an ad may lead to lower conversions, or vice versa. By optimizing for CPI, you ensure that neither one of those success metrics is compromised during the optimization process. Like all success metrics based around conversions, however, a large sample is needed to prove a winner statistically.

Setting Up Your Tests

Once you have a hypothesis and an understanding of how to determine your results, it is important to determine the type of test you want to run.

Prior to creating your test ads, you need to decide on an A/B or Multivariate Test. This choice will determine the creative variations to implement in the test. Before choosing which type of test to run, it is important to understand their differences and what types of insights each could yield.

An A/B test, also known as univariate, is structured so only one variable is changed. Different versions of the same variable may be tested, but the changes are still only being applied to one variable. For instance, an A/B might test one ad with the CTA of *"30 day free trial"* and the other with *"try it for free"*.

Assuming a third variation of the CTA, *"free trial now"*, is added to the test. Does it then become a Multivariate? The answer is no. A third variation makes the test an A/B/C test and a fourth variation makes it an A/B/C/D test, all which fall under the A/B testing umbrella.
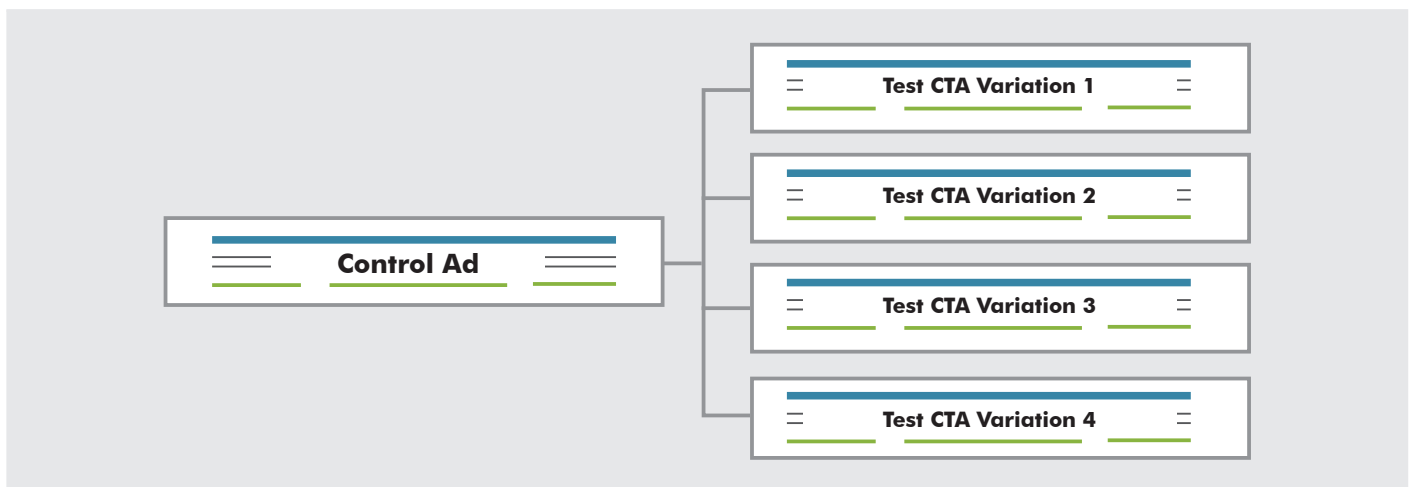
In multivariate testing, on the other hand, two or more variables that are nominally independent of one another are being tested. For instance, the test ad in this case would have a different headline and Call-to-Action from the control ad.
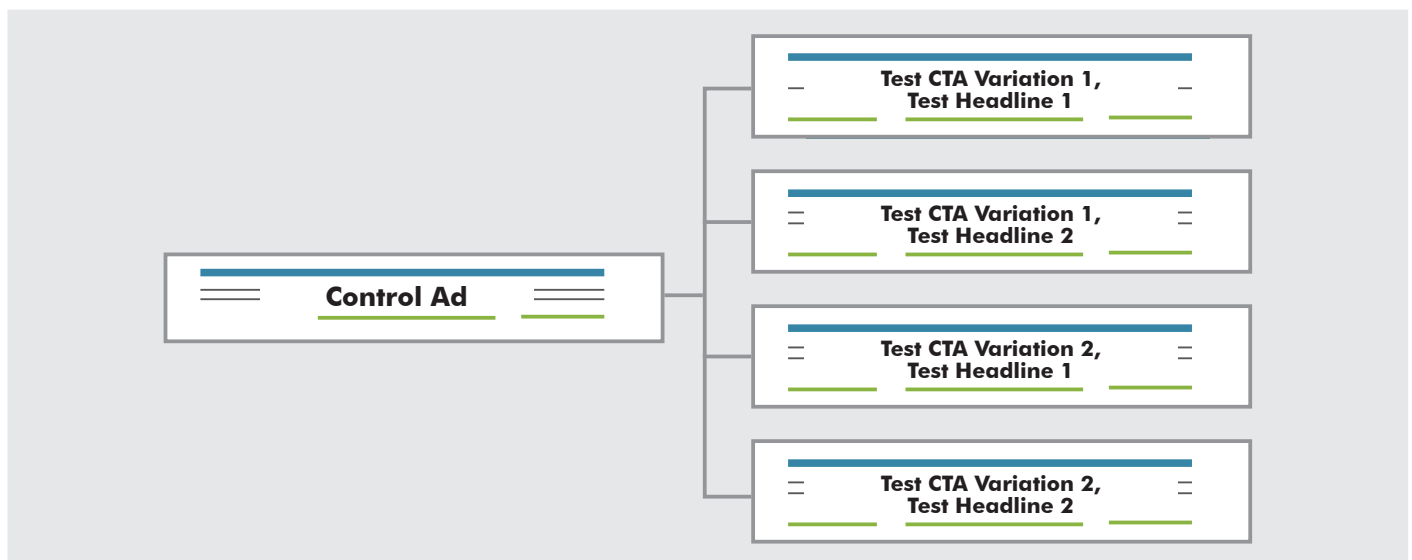
When it comes to making a choice between A/B and Multivariate tests, it is important to note that A/B testing is simpler, allows for more control, and the results are easier to interpret. It is often thought of as the best way to pick the low-hanging conversion fruit.

	A/B Testing	**Multivariate Testing**
+ Pros	• Simpler • Allows for more control • Easier to isolate influential variable • Results are easier to interpret	• Often yields high-impact results
- Cons	• May yield lower-impact results	• Difficult to isolate element that leads to high-impact results • Easy to get carried away with wanting to test all combinations

Multivariate testing can have a higher impact, but it is more challenging to isolate the variable that led to a certain outcome. In the previous hypothetical example where the headline and CTA were changed, had the test yielded positive or negative results, it would be difficult to determine if the cause was the new CTA or Headline. Finally, with Multivariate testing, it is often easy to get carried away with wanting to test every possible combination of variables and variations.

The bottom line is that finding highly influential variations is generally much more valuable to an advertiser than testing all the potential combinations. They are also easier to isolate and test or apply across other areas of the account.

Simply stated, the question then becomes: *"Do I want maximum impact or do I want to clearly understand impact of individual elements?"*

A good starting point would be A/B testing of broad differences to find out what resonates with searchers at different stages in the funnel. Once that is determined, more sophisticated Multivariate tests can be implemented to refine ads and drive incremental improvement.

Testing Tools to Use

When conducting ad tests on Google, advertisers have a number of options which impact the way the test ads rotate and accumulate volume. These settings can impact several factors of your test, and each has its benefits and challenges.

Optimize for Clicks

Google's 'Optimize for Clicks' option is the default setting in AdWords and if selected, Google will show the ad with the highest click through rate more frequently. While this will likely get you the most clicks possible in test ad groups, it will also come at the expense of your insights. Google is quick to make decisions on which of your ads has better CTR and it is unlikely that all of your ads will get enough traffic to determine which actually does best from an overall performance standpoint. Additionally, as seasons change, different copy tends to perform better, and Google will rarely resurface a 'losing' ad to see if it might 'win' at any point in the future.

Optimize for Conversions

'Optimize for Conversions' has a similar approach to 'Optimize for Clicks', however, Google will show the ad that generated the most conversions more frequently. This setting requires that the advertiser allow Google pixels on their site to track conversion activity. Recent studies conducted by Google have shown that campaigns switched from 'Optimize for Clicks' to 'Optimize for Conversions' saw an average of 5% increase in conversions. However, similar to 'Optimize for Clicks', you should realize that on this setting Google will be quick to choose the ad that generates the most conversions, so some of your ads will not get as much of a chance to prove themselves worthy.

Equal Rotation (Round Robin)

Google's "Rotate" setting is preferable for reliable ad testing, since every ad in a given ad group will receive a nearly equivalent number of impressions in similar positions. With this setting you can do a more apples to apples comparison of results and know with confidence that when you select the winning ad, it is truly the winner as defined by your success metrics. The biggest downside is that you may sacrifice clicks or conversions running traffic against lower performing ads as you wait for your test to complete.
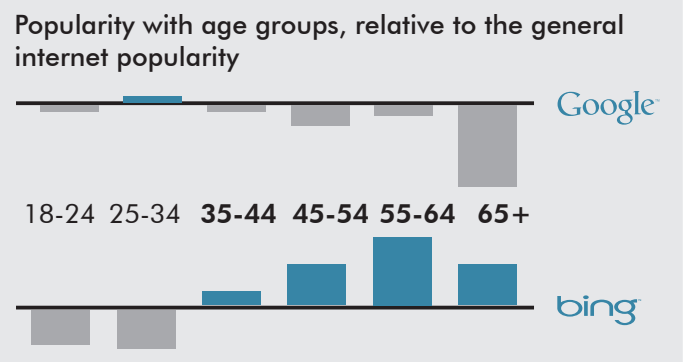
AdWords Campaign Experiments (ACE)

ACE is a newer testing tool in the AdWords interface that allows you to see how changes to many different elements of your campaign, including keywords, bids, ad groups, ads and placements, affect overall performance. It's a powerful and underutilized tool. For many, the main advantage of using ACE to test ad copy is risk mitigation. It provides advertisers the ability to apply the experiment settings to a number of campaigns across their AdWords account, and within each campaign advertisers can determine to what percentage of the traffic the experimental changes should be applied. Using ACE requires setting a percentage that you are comfortable with, being patient as the data accrues, monitoring your results and making decisions based on statistical significance.

Bing Ads

Bing is a different traffic source than Google so although the two engines have their similarities, different tests are often needed, even for similar keywords and ads. Like ads in Google, Bing ads are engineered to be triggered based on the user's search query. However, the two engines have audience and ad marketplace differences. As such, it is smart to continue to test your ads on Bing, even if you've gotten clear results on Google.

Some challenges to ad testing on Bing: the ad rotation options are unavailable in the Bing Ads Desktop Tool and as such they much be added to campaigns and ad groups manually, one-by-one. Additionally, a significant percentage of traffic on Bing comes from network partners, which has the potential to confound or obscure results. Age differences across Google and Bing are just one signal of the different audiences, requiring unique ad testing for each engine.[4]

Popularity with age groups, relative to the general internet popularity

Google

18-24 25-34 **35-44 45-54 55-64 65+**

bing

Knowing When Your Test is Done

Statistical Significance

Statistical significance is a reflection of whether or not the relationships observed between the variables in a test do in fact impact the results, or if the differences in the results have occurred due to chance. A simpler way to understand the concept of statistical significance is to think of it as a way of revealing to which degree the test results are determined by the variations you created.

One of the most common questions when it comes to testing ad copy is how long you need to run to constitute a valid test. This is going to vary greatly based on traffic levels in your test group and your specific success metrics. There is no one-size fits all impression count that determines when a test is over since CTR and CR vary greatly from advertiser to advertiser. However, statistical significance helps level the playing field. Most scientific experiments use a statistical significance of 95% in a standard t-test to determine if a test population is distinct from a control.

It is important to note that while other key metrics may reach statistical significance during your test, you must continue to focus on the statistical significance of your original success metrics before ending a test. If a metric has not changed by a statistically significant amount, then it is safe to assume that it was not affected by your experiment, and the differences which occurred were only experimental noise.

There are helpful tools and templates that you can use to help determine statistical significance, such as Chad Summerhill's simple, but highly effective statistical validity tool for ad tests.

Statistical Significance Validity Tool[5]

Confidence Level 95.0% ▼

	Impressions	Clicks	Conv
Control	49110	2204	874
Test Ad	35959	2243	906

	CTR	CR	Imp2Conv
Control	4.49%	39.7%	1.8%
Test Ad	6.24%	40.4%	2.5%

		Needed Imp.	2533
Optimize for CTR	Actual Diff.	Needed Diff.	Lift
	1.755%	0.310%	39.1%

		Needed Clicks	67878
Optimize for CR	Actual Diff.	Needed Diff.	Lift
	0.737%	2.880%	1.9%

Normalization of Results

As tests run and data starts to accrue, it is not unusual to see a dip in conversions. Some may find this alarming and stop their tests. In reality, data take time to normalize, and so it is absolutely crucial that you remain patient and not act on results prematurely. Additionally, you need keep in mind known trends such as your sales or conversions being higher on weekdays, other seasonality, traffic fluctuations, and the dangers of using December data.

The following two charts highlight are illustrations that highlight how long it could take for ad test results to normalize. The top chart is for a test where the success metric is CTR – notice that the ads in the test arrive at their long term performance reasonably quickly. In the bottom example, where success is defined by conversion rate, it takes weeks before the two ads arrive at their long term performance profile.
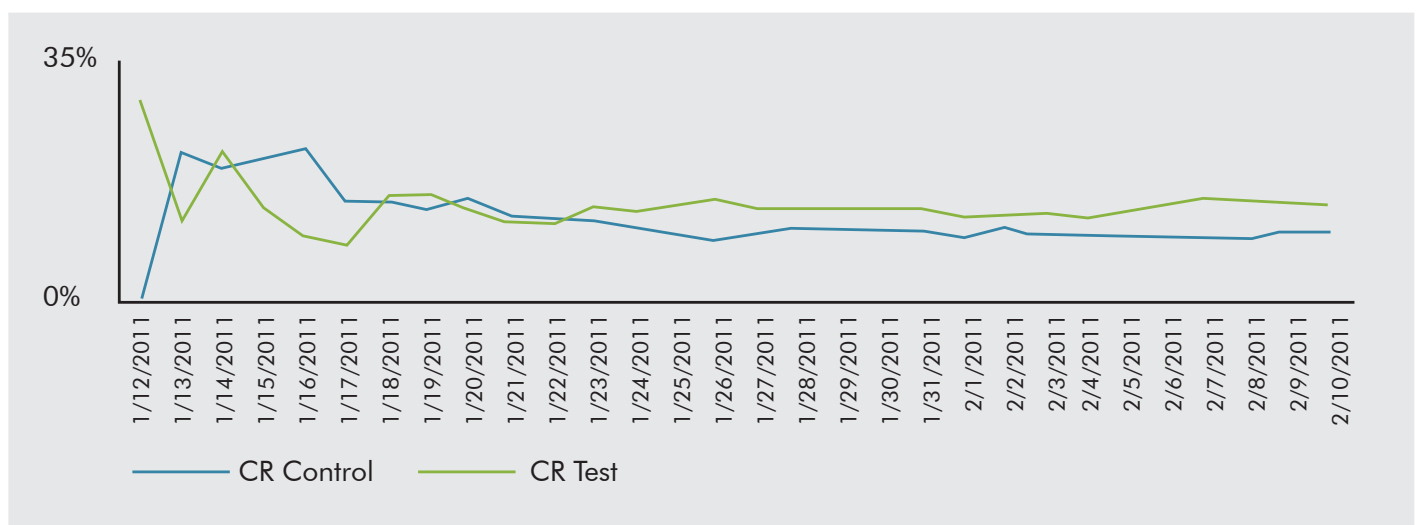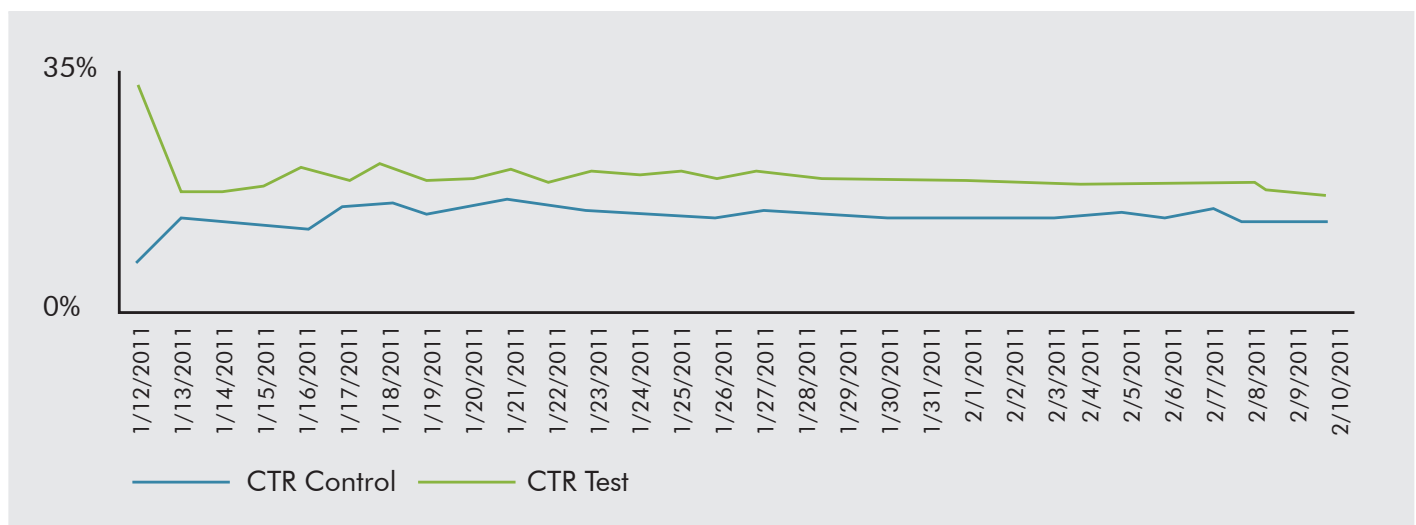
Leveraging
Test Results

After a test is concluded, a winner is declared and before acting on the data, you should refer back to your hypothesis to ensure that the test results satisfy the intended test.

The next step is to understand which ad is better based on your original goals. Because it is not uncommon for more than one metric to change, it is often difficult to choose a "real" winner. Would an ad that increased CTR but lowered conversions, or vice versa, be considered a winner or a loser? In these instances, it may seem like there is no best, so the choice will depend on the overarching business purpose of the test and which metrics align with those goals. The following are guidelines to choose winning ads based on goals:

If your goal is:	Choose ad with:
Customer Acquisition	Highest Conversions
More Profit	Highest Profit per Impression
Higher Traffic Volume	Highest Click Through Rate
Lower Lead Costs	Lowest Cost per Conversion

After choosing the top performing ad based on success metrics, that ad should become the new control, underperforming ads should be paused (not deleted!), and new ad copy created and tested against the new control.

Winning ads should still be monitored long after tests are concluded. A month or so of monitoring CTR, position, conversions, and impressions of winners will ensure that previous test results were not affected by things such as seasonality or irregular traffic patterns.

Insights drawn from tests, regardless of whether or not they were successful, should not only be about which variation of ad copy was better than the other. Smart advertisers look beyond just the ad unit to try and understand which messages within the ad resonate and why certain ads perform the way they do. They leverage these insights to gain a better understanding about their target market, and even test successful messages across various other online and offline channels.

Testing Across Ad Groups

Advertisers of all sizes can benefit from ad optimization results by finding ways to apply successful tests across their accounts. Because insights should never be generalized it is important to semantically group ones with similar or common themes or intents when doing tests.

For example, a specialty retailer may find that all SKU level ad groups represent searchers who are ready to buy and are beyond their research phase. Now they can test ad elements that worked best in one SKU ad group in other SKU ad groups with some reason to believe that they may see similar beneficial performance.

Low traffic accounts can benefit from ad testing even though most ad groups have limited traffic volumes. A recommended method includes writing two different ad themes and testing them across a variety of ad groups with similar intents. By coordinating the test across multiple ad groups, advertisers can still see the impact each theme has on success metrics.

A case study in testing across ad groups with similar intent

Testing across multiple handbag ad groups enabled BagBorrowOrSteal to identify that color messaging resonates with luxury shoppers.

Louis Vuitton Handbags Stylish **LV** Handbags in 18 Colors like **Turquoise** and **Red**. BagBorrowOrSteal.com/Louis-Vuitton	**Brand + Color**

Rent Prada Handbags **Prada** Handbags from **$34**/week. Join and Save! BagBorrowOrSteal.com/Prada	**Brand + Price**

Developing an Ongoing Testing Framework

Ad optimization should be an ongoing process, and if you don't put a repeatable structure around it, chances are it will fall off your radar screen at some point.

To that end the following optimization cycle provides a suggested timeline and the recommended actions to take to ensure that the best possible ad is running at any given point in time. However, the ultimate framework you use may be different than this, depending on your industry, seasonality, account traffic volume, sales cycle, etc.

The most important thing is that there is some systematic process around how you do this. Additionally, there is a worksheet provided in the Appendix to help you keep track of your testing efforts.

Optimization Cycle	Recommended Actions
Daily	Refresh **dynamic data** (e.g. price, inventory) Check **status of ad tests** currently running
Weekly	Introduce **new ad themes** into rotation Wrap existing tests that have **significance** Review any **new offers** for introduction into ads
Monthly	Check **competitive landscape** to inform new ads Review **customer intent shifts** to drive structure changes
Seasonal	Full **inventory refresh** to capture updates Introduce **seasonal promotions** into rotation
Annual	Test new or proposed **brand messaging**

Conclusion

Any strong PPC account manager would tell you that there is no "best" ad, just a little better than the previous one. Optimization is a continuous process where you should constantly try to outperform your current 'winner.' However, even if test ads perform worse than the control, there's knowledge to be gained and insight to be leveraged.

Seasoned advertisers who test and optimize on a regular basis know better than to be discouraged by test results. In the beginning, it is not uncommon to see dramatic improvements and changes, but over time, they start to be a little less drastic and a lot more incremental. By systematically testing and optimizing using the techniques laid out in this whitepaper, you can drive huge improvements to their bottom line and improve the overall quality of their PPC accounts.

Search marketing has historically been the best performing channel in the marketing mix, but it is getting increasingly more competitive. It is time to walk away from keyword wars and start driving campaign performance through active ad optimization. By continuing to focus on what makes your brand unique and engaging to consumers, you will create a sustainable competitive differentiator.

1. Geddes, Brad. Advanced Google AdWords, Second Edition. 2012
2. Healy, R. Text Ad Testing Master's Guide Q&A with 18 PPC Experts,
   Retrieved December 2012, from http://www.ryanhealy.com/reports/Text-Ad-Testing-Masters-Guide.pdf
3. Eisenberg, B. The Conversion Trinity: The 3 Step Magic Formula to Increase Click Throughs & Conversions. Retrieved December 2012,
   from http://www.bryaneisenberg.com/the-conversion-trinity-the-3-step-magic-formula-to-increase-click-throughs-conversions/
4. Alexa.com, Retrieved December 2012Alex.com, Retrieved December 2012
5. Source: Summerhill, C. Free Excel Spreadsheet for PPC Text Ad Testing for Statistical Significance.
   Retrieved December 2012, from http://www.chadsummerhill.com/ppc-text-ad-test-statistical-validity/
6. "Importance of Google Text Ad Optimization". Retrieved December 2012,
   from http://www.clickequations.com/blog/2011/05/importance-of-google-text-ad-optimization/

Appendix

Ad testing worksheet

1. Define your goal	
My goal is:	
I will measure this by:	
My success metric(s) will be:	
I am testing:	
I am using the following type of test:	
My variables are:	
My variations are:	
2. The hypothesis	
My hypothesis is:	
My hypothesis is based on:	
3. The control variable/variation	
My control variable is:	
My control variation is:	
My test variable is:	
My test variation is:	
4. The test	
I will run my test for a duration of:	
I will test to a statistical significance of:	
5. Results	
The following metrics improved:	
The following metrics worsened:	
What I learned is:	
I will use these findings to:	

About DataPop

DataPop's creative optimization technology develops the optimal messaging and most engaging ad for every consumer. Leveraging the technology allows advertisers to improve performance in the short term, and gives them insights to replicate that success over time. DataPop clients, on average, have seen sales increases of 40% while also experiencing CPA declines of 15%.

DataPop

e: info@DataPop.com
p: 323-302-4987
w: www.datapop.com
t: Twitter.com/DataPop

Dave Schwartz

e: Dave@DataPop.com
p: 310-409-8100

Frank Lee

e: Frank@DataPop.com
p: 213-700-2263