5 Deadly Ad Testing Mistakes You Probably Made This Week

by Mark Joyner
www.MarkJoyner.name

If you are testing your advertising at all, you should pat yourself on the back. You're in good company. It's been estimated that less than 5% of all business owners track the results of their advertising at all.

The problem is, even if you do track your results, chances are you are making some rather critical mistakes that will render your testing almost useless.

The good news is that these errors are easy to correct once you know how to identify them.

1. The "Monday-Tuesday Switcharoo" Error

I was talking to a friend of mine who just came back from a marketing seminar and he was full of excitement.

"Mark did you know that if I add this widget to my site I will increase my orders by 419%?"

Trying to stifle my wincing (he obviously didn't know the "Your Site = My Site" error, but I get ahead of myself), I inquired further.

I said, "Really? If I put that widget on my site it will increase my sales, huh? How do you know this?"

"Well, the guy at the seminar was really smart. Here's what he did. He ran his website with the widget and got 18 sales. The next day, he ran his site without it and got only 10!"

What my friend failed to realize (which apparently this expert failed to realize as well, so I can't fault him) was that Monday does not equal Tuesday.

Example: let's say you run version A of your website on September 10th and version B of your website on September 11th.

Oh yeah, the year is 2001.

Without looking at your copy, I could almost guarantee that version A was the winner. Why? Well something pretty major happened on September 11th 2001 that froze the sales of just about everything all over the world.

This is an extreme example, but every day is the same to a lesser degree. Things happen from day to day, totally outside of your control, which could affect your sales. If you fail to account for that your testing is absolutely meaningless.

In order for any scientific testing to be valid, you have to "isolate your variables." Market testing is, in fact, a scientific practice albeit a rather loose one.

What the heck does all that mean? Well, simply put: you need to make really sure the question you're asking is truly being answered by your test.

If your question is: "Is A better than B in generating sales?" then you need to "control" anything in your test which may *also* affect sales.

If you run version A on Monday and B on Tuesday, you're also asking another question at the same time: "Is Monday better than Tuesday for sales?"

Two questions at once = bad marketing mojo.

It just confuses things and removes any and all hope of getting a meaningful answer. At the end of such a test you have "information," but there is no way of knowing what it actually means.

If this is confusing, here's an example that will make it really clear.

Imagine you have two batches of banana seeds. You want to see which of these batches of seeds will grow trees that render the most fruit.

Simple question, right?

So, you take one batch and you plant some of them in your backyard.

You take the other batch and you mail it off to your friend in Paraguay and ask him to plant them.

If you compare the number of bananas from your trees to the number of bananas on your friend's trees, will you know which lot is better?

You've probably already figured out that the answer is "no." Why? Because it could very well be possible that Paraguay is a better (or worse) place to grow bananas than your backyard.

You didn't ask just one question. You didn't "isolate your variables."

You really asked several questions at once:

"Is Paraguay better than my backyard for growing bananas?"
"Is my friend a better gardener than I?"
"Is the soil in my friend's backyard better than mine?"
And so on …

If you really wanted to test this well, you would have needed to use the two different batches of seeds in the same **soil under** the **same conditions**.

Any scientific question must be asked in this manner or the answer you get is a meaningless one.

The way to overcome this on the Internet is to use a split testing script that rotates between A and B in real time.

That is, you do your best to evenly distribute the two versions across your viewers so as to eliminate any other possibilities that could affect your results.

In the offline world, people do "A/B split testing" all the time by:

- Sending out Direct Mail Letter A to 50% of their list and Letter B to the other 50% at the same time and counting the results. (Oh yeah – A and B need to be randomized – that is, mix up the list randomly, split it in two, and then send.)

- Asking a magazine to show Ad Version A to a random 50% of it's readers and Ad Version B to the other 50% of it's readers.

- And so on …

You can run tests as the marketing consultant who misled my friend did, but the results you get will mean absolutely nothing.

Ask one question at a time and you're more likely to get a useable result.

To do this online, you need to use some form of "split run testing" script. They are readily available online and I can't recommend any one solution over another.

2. The "Your Site = My Site" Error

Let's look at that example again.

Joe Guru adds a widget to his site and it increases his sales. Let's imagine for a moment that he actually performed his test properly and his observation was in fact meaningful.

Does this necessarily mean that you will be able to add this widget to your site and increase your sales as well?

Unfortunately not. If learning marketing lessons were so easy I would own a small island the size of Australia by now.

See, it's possible that while this change affected his site positively it may not have the same affect on your site.

Let me give you an example.

Let's say Joe Barber has a website for his barber shop. He let's people book appointments via his website and he measures his success by the number of appointments he books each day.

One day he decides to test the response of his site by adding a picture of himself to his website. He does good A/B split run testing and gathers enough data to make a valid observation.

He determines that by adding his picture to his site that he does indeed increase his response rate.

He tells you this story and you make an assumption:

"Adding Joe's picture to a website increases response!"

So, you add Joe's picture to your "How to Throw a Bridal Shower" website and you don't test it, because you assume it will increase your response rate.

Wait a minute …

Did you do a double-take there? You should have.

It would be incredibly stupid to put *Joe's* picture on *your* website, right? Especially if Joe is a big greasy dude who caters to other big greasy dudes needing haircuts. His clientele may be warmed by his picture ("Hey, he's a big greasy dude, too! I don't have to feel threatened when I walk in to get my hair cut.") but the ladies visiting your site wanting to plan a bridal shower may not be so impressed.

OK, that one's obvious. Let's look at it another way and see if we can still gather anything meaningful from Joe's experiment.

Let's say we look at the experiment again and say, "This must mean that adding a picture of myself to my website will increase my response."

That's a smarter conclusion than the first one. Is it valid?

Think for a minute …

Let's see if you were right.

The answer is "no."

Why? Let's take a look.

Imagine for a minute that you're a 40 year old man who decided that catering to the Bridal Shower niche market would make you some extra money. May not be a bad idea. People are getting married every day and since throwing a bridal shower is not an everyday occurrence, it stands to reason that the ladies throwing these parties may need a bit of help.

However, do you think the young ladies going to the site will respond well to your picture? (You're a 40 year old man, remember?)

I would tend to say no …

But you know what? You really don't know. You'd have to test this for yourself on your site under your own conditions.

Maybe one picture would work better than another.

Maybe a stock photo would.

Maybe no picture at all is best.

The point is, just because something works on someone else's site doesn't mean it will work on yours.

Does this mean that we can never draw any conclusions from the marketing of others? Definitely not. If we didn't, there would simply be no progress in the field of marketing at all.

What works on someone else's site may in fact work on your site. You don't know.

Now, if something works on 1,000 sites out of 1,100 site that have tested it, you have a pretty strong observation in your hands (assuming the tests were carried out in a proper scientific way).

I've released observations from my own marketing to the public in the past in hopes of improving the results of the readers. However, I always did so under two conditions:

a. I tested it extensively under varying conditions in order to increase my chances that the observation was in fact meaningful and likely to be beneficial to the marketing of others.

b. I always included a caveat: this worked on my sites – it may not work on yours. Even though I've proven it fairly rigorously, you have to test it out yourself before you can be sure if it will help you or not.

If a marketing observation has been made on 1,000 sites all in the same niche, the same observation may not hold on another.

Another example: observations made in the field of shopping cart marketing may not have any positive impact on direst response salesletters.

And so on …

Does that make sense?

Take what you are told with a grain of salt. With the knowledge you just learned, you can ask those giving you marketing advice meaningful questions. Demand to know their test conditions. If their conditions are invalid, call them on it. It will improve their game and will help those to whom they give their advice.

You can now arm yourself with the most powerful marketing weapon of all: a sharp and observant mind that understand the true meaning of marketing experiments.

3. The "Gestalt" Error

Imagine you have a direct response website and you test two elements:

Your Headline (and)
Your Order Page

Good choices! Those two elements are more likely to affect your response rate than anything else.

Now, let's say you were running a multivariate split run test for both of those elements at the same time.

"Multivariate" is just a highfalutin way of saying, "We're testing more than one variable at the same time."

You look at your data and you determine that Headline A is the best between Headlines A and B. So far so good. Then you determine that among Order Pages C and D, D is the best.

Great!

So now, all you have to do is put Headline A up and Order page D up and – Bob's-Yer-Uncle! – you've got a winner, right?

Sadly, no.

Headline A may have been the winner overall, but there is a possibility that it performs poorly when *combined* with Order Page D.

Back at Aesop I coined two phrases to refer to this phenomenon: "Gestalt" and "In-a-Box."

Gestalt Testing

The phrase "gestalt" is a word we borrow from German meaning:

> **n.** *A physical, biological, psychological, or symbolic configuration or pattern of elements so unified as a whole that its properties cannot be derived from a simple summation of its parts.*

Was that total nonsense to you? Don't worry. The important part is simple: it's something that can not be explained as a sum of it's parts.

In Gestalt Testing we take into account that testing data is ***not*** additive. That is, we can not say that combining two winning elements will necessarily render an additive best result. We theorize that the "Gestalt" of (or the combination of) different elements must be looked at together.

Headline A may very well be the best.

Order Page D may very well be the best.

However, you can not be at all certain that the two combined will perform well.

They might … But you can't know for sure.

Gestalt Testing traditionally would take much more time to run. That is, we would have to gather significant data not on each part by itself, but on how it combines with others.

If you look at Headline A, you have to look at it twice: once with Order Page C and once again with Order Page D. This means you have to gather twice as much data.

However, the results are a far better indicator of the truth if you're testing many variables in combination.

In-a-Box Testing

This is the opposite of Gestalt Testing. It totally disregards the Gestalt and looks at the variables in isolation.

If you are testing multiple variables looking at the In-a-Box data could be meaningless. However, if you are limited to the amount of information you can gather, you may be forced to fall back on the In-a-Box data and that's better than nothing.

……..

A multivariate system called "Taguchi" which has come in to some fame these days, unfortunately fails to take the Gestalt Error into account. After studying it in depth I realized that the way Taguchi makes it's judgments is by assuming that all of the variables it tests are additive.

Now, this may have it's uses at times (especially in something like direct mail where running a full multivariate test would require many expensive print runs), but if you're using Taguchi, don't kid yourself into thinking it's not addressing the Gestalt issue.

Taguchi works by running a relatively small number of tests with a carefully selected grouping of variables. It then predicts which combination of these variables will perform the best.

The problem is, the algorithm doesn't actually test the combination itself. It predicts which combination will work best, and frankly the algorithm has no way of actually knowing this.

This algorithm works in the manufacturing industry, for which Dr. Taguchi designed it, because in engineering variables *are* additive.

I went so far as to call up a "Taguchi Expert" to ask him about this, and only got a rather uninformed response. He didn't quite understand my questions, quite frankly.

Even still they wanted to charge me $40,000 for a consultation to teach me their system! I wanted to get a fair idea of what they'd teach me before I laid out that kind of cash.

I asked, "How can Taguchi know if your recommended combination of variables will perform well together? Is the algorithm psychic?"

"Oh, well Mark, it's just the magic of Taguchi."

That was the end of a one hour conversation and I politely informed him that he should look into selling used cards and hung up.

"Magic of Taguchi, huh?" I will call him next time my buddy Jack needs some magic beans.

I'm sure this section is going to piss off some Taguchi fans, but I'm sorry. Look into it. It plain assumes that variables are additive. Do some testing yourself and you'll find that a significant amount of the time they are not.

And no disrespect is meant to Dr. Taguchi. His algorithm obviously revolutionized the manufacturing industry, but I believe it has this one major flaw when applied to marketing (which was never his intention).

Don't take my word for it, though. Test it out yourself. Just the opinion of one marketer who has looked into it rather deeply.

4. The "Law of 40" Error

This is also sometimes called "The 25 Response Rule" or "The 40 Response Rule" and so on … depending on who you talk to.

Unfortunately, even some of the best marketers I know still use this rule. I'm sorry to say after researching high and low I can't find any scientific basis for it.

What am I referring to?

A little background …

In any scientific testing, you need a way to measure how "confident" you are that the results you have are meaningful.

For example …

Let's say you wanted to send out an email campaign to your list. Before you mail it out to your entire list, you decide to test a few different versions of your message to a smaller chunk of your list first.

The idea here is that you are more likely to select a good campaign if select the best of several options. If you try only one, there is no telling how well it will perform.

Why is this a wonderful thing?

Well, let's imagine that you have 5 different email campaigns and the results of each is as follows:

Campaign 1: 5.1 Sales for Every 1000 Mails Sent
Campaign 2: 2.3 Sales for Every 1000 Mails Sent
Campaign 3: 0.6 Sales for Every 1000 Mails Sent

Campaign 4: 1.2 Sales for Every 1000 Mails Sent
Campaign 5: 4.5 Sales for Every 1000 Mails Sent

Obviously these are averages.

Now, if the trend you observed continues, selecting campaign 1 will obviously have a huge impact on your results. 10 times great than the results from campaign 3 to be precise!

If that's the difference between 50 and 500 sales, that's a pretty stunning result.

The big question is, how do we know that this trend will in fact continue when you mail out the campaign to your entire list?

Imagine that your list is actually 420,000 in size. You sent each of your tests to a random sampling of 4,000 members.

After sending out that many emails, will it accurately predict how well the rest of the list will perform?

4,000 seems like a lot, right? You'd think you have enough information to make an accurate prediction, but the answer may surprise you.

Let's get back to science for a minute …

There are various tools in the field of Statistics that will help you to determine this, but most marketers do not use them at all.

Statistics provides a number of ways you can measure the "confidence" of your test results. That is, you can enter information about your test in to a mathematical formula and you will be able to calculate, to a very accurate degree, exactly how confident you can be about your predictions.

That is, you can know, within a certain margin of error, whether or not the results of one of your tests will accurately predict what will happen if sent to the rest of your list. (Or in the case of a website, if your test is an accurate prediction of how people will behave looking at the same ad copy in the future.)

Sadly, this concept is foreign to most of the marketers I know. I can't blame them, though. They are engrossed in writing ad copy and concocting clever ways to make sales and god bless them for it. That's where their minds should be.

However, as a group we need to clean up our act here a bit and get more precise.

For the most part, marketers look at their test, select whichever campaign has the best response, and roll out their marketing with what appears to be the winner.

That's better than nothing, but whether or not you can truly be "confident" that you're really sending out your best marketing to your entire list is unknown. If you haven't measured the confidence of your predictions, you could very well be picking a losing campaign rather than a winning one.

You just wouldn't know.

Some marketers intuitively understand this concept, so they intelligently ask the question: "How can I be confident that my test is an accurate prediction of the future?"

It's a great question – one of the ultimate questions in marketing in fact – but the direct response marketing world does not have an accurate answer for it.

The standard response is, "Run your test until you have 20 responses (or 25 responses or 40 responses depending on who you're talking to) and you will have enough information to make a decision."

When I first heard this I scratched my head saying "this just can't be right." Everything I knew about statistics said that this was a meaningless statement.

I have to tell you after several years of searching and researching, I can't find a single piece of scientific information to reliably justify the use of "the law of 40" in your marketing.

Here's why.

To measure confidence, you need to take into account several things:

A. Your Sample Size
B. Your Response Rate
C. The Number of Responses

To blindly say, "Test until you get 40 responses" is simply unjustifiable because "40 responses" means absolutely nothing in a vacuum. The number of responses required to have any level of confidence is totally dependent on your sample size and response rate as well.

Now, the "law of 40" is again better than nothing. At least it disciplines you to get a fair number of responses before making a decision.

However, don't fool yourself into thinking that you can be certain you're selecting your best marketing campaign when using this rule.

Let me just state that no offense is meant to the folks who are teaching this. Some of them are my near and dear friends and they are just passing along the knowledge that was

passed on to them over time. This is the way we learn and it just so happens that sometimes we learn things that are wrong.

I will challenge my colleagues, though, to ponder some of the conclusions they have made on testing conducted using the Law of 40 as a measure …

What's really interesting is that many of the folks using the Law of 40 are getting tremendous results. Now, this tells me one of several possibilities (or a combination thereof) …

a. The Law of 40 Works Sometimes

   It stands to reason that it would … If the other elements line up properly it will at times be an accurate measure.

b. You May Not Be Selecting Your Best Marketing, But You're Making an Improvement

   This stands to reason, too. It may prevent you from picking a total dog of a campaign *sometimes*.

c. Maybe There is More to This Than What We Can Measure …

   I won't elaborate on that and will leave it to your imagination.

…

It should be emphasized at this point that even traditional confidence measurements can be troublesome. The amount of data required to accurately predict an email campaign, for example, is astronomically large. Most people simply don't have lists large enough to make accurate predictions.

However, there is an alternative that, as I write this, only me and one other person on the planet know about right now. Check out this page to learn about "**Mike Chen's Secret Weapon**." It will only be available between March 18th and March 30th 2005.

5. The "Paralysis by Analysis" Error

A good friend of mine is running magazine ads for a cosmetic product and doing quite well. He hit a very solid ROI for one of his campaigns and by all rights he should have been very happy. For every $1 he was spending he was making about $1.50 back. Not bad at all!

At this point, he should have begun rolling out his campaign more aggressively.

He suddenly got it in his mind that he should begin testing price. He started a new test campaign at a higher price and wanted to compare it to his old price. Not a bad move. It's very possible that he could even increase sales with a price increase, so it's a wise test. What he did after he got that notion, though, was not so wise…

He gave his test ad campaign to his media buyer and asked them to run it as quickly as possible. Then he waited …

And waited …

And waited …

I met up with him for lunch and asked him how his campaigns were doing and he said, "Oh man, we're stuck. I can't do a thing until I get the results of that price test back and the media buyer keeps stringing me along."

I paused and then asked him …

"Are you getting an ROI with your existing campaign?"

"Yeah, we're getting about 150% ROI across the board easily."

I looked at him and said, "I want you to listen to me very very carefully. Don't answer this without really thinking hard. You're going to want to respond quickly, but don't. Answer me this: why don't you just roll out hard with the campaign that's already working?"

He paused for a minute and then his face lit up.

We quickly ended our lunch so he could go back to the office and start ramping up his campaigns with the price he already knew was making him money. If he finds out later that the other price is better, no problem. He just carries on with the new price. There's no reason to stop the marketing you already have in place that is rendering an ROI.

This may seem obvious looking from the outside, but when we're the ones making these decisions it's not easy at all.

We sometimes stop marketing because we think a better campaign is on the way. If you have marketing that is working, you should push ahead with everything you've got.

Depending on the risks involved, I'd even sometimes say that you should push ahead with ***something*** even if it's totally untested. Testing is a wonderful thing, but if it becomes an impediment to action I'd be tempted to tell you to stop testing.

On the scale of things, action is far more important than testing. Some businesses can make it without testing (again, depending on the risks, the advertising medium, and the market) but no businesses will *ever* make it without action.

A foolish decision today is better than a brilliant decision next year (or next *never*).

All the best,

Mark

Mark Joyner
www.MarkJoyner.name

P.S. On 18 March 2005 Mike Chen is going to release **his secret weapon** to the public for only two weeks. As you may know, I consulted for Mike on his projects InstantBuzz and Make Your Own Software. Through the course of that work we did together, Mike created something absolutely extraordinary. If you're at all serious about achieving any success with your business, you'll thank me for telling you about this if you're one of the lucky few who act. I dare say it's going to give a huge advantage to anyone who gets it.