

3-Part Whitepaper Series: The Top Objections to Testing (and Why They're Wrong!)
By Brian Tvenstrup

Part 1: The Importance of Testing

What is testing and why is it important to your business?

Many business executives don't understand the importance of using well-designed tests to help them improve their business results. In some cases, this is a result of a dominant organizational or industry culture in which the role of empirical data analysis has been historically weak outside of very limited contexts. Other times, it is the personal bias of a decision-maker that inclines towards intuition and judgment and away from quantitative analysis. In rare cases, executives will have been exposed to abuses of testing in prior experiences and fear "analysis paralysis" or even worse, out-of-control expenditures of time and resources in the name of testing that fail to generate any valuable output for the business.

In spite of widespread use in certain industries, testing as a standard tool to enhance decision-making across the entire enterprise has been neglected in others. This can even be true within the same industry, where testing is used extensively in one setting but only sporadically in other adjacent activities. For example, in the pharmaceutical industry, controlled testing is a hallmark of clinical trials and medical impact studies, but it is routinely ignored in marketing contexts.

In this whitepaper series, we will explore the top reasons why business leaders should not take shortcuts when it comes to using empirical methods of testing. When designed and deployed properly, testing is a business analytics strategy that should lead to significant improvements in decision-making across the business. But for the purposes of this discussion, we will take a skeptical view of testing and explore some of the common objections to testing as a lens through which to examine its benefits. So let's consider the top reason that executives believe that they don't need to invest in testing for their business decisions.

Objection #1. I don't need tests to measure the impact of my activities. I can use "Business As Usual" (BAU) results and changes over time to estimate that directly.

This is probably both the most common and the most foundational objection to the use of testing in business decision-making. But there is a logical fallacy at the heart of this particular objection, and it is important to bring it out into the open. The key point is that no one is actually able to observe simultaneous alternative states of the universe in advance. Anyone who had such powers would not be wasting their time working in the world of business! They would be rich and powerful beyond measure and we can only hope that they would use their powers for good.

What testing allows you to do is observe simultaneous alternative states of the universe, but retrospectively. If designed properly, testing can answer questions like "what is the difference if I

send this particular communication in green envelopes versus red envelopes?” If you have multiple possible tactics, each of which is a plausibly good approach, then you should test them in head-to-head comparisons in which the population of interest is divided randomly into subsets, with some groups receiving one tactic and other groups receiving a different tactic. This kind of head-to-head testing is commonly called “A/B testing” when there are two tactics being tested (one arbitrarily called “A” and the other “B”). In a somewhat more evocative metaphor, it can also be called “champion-challenger testing,” which is usually the case if one of the tactics is considered to be the preferred treatment based on prior experience or results (the “champion”), and the other is considered a newcomer (the “challenger”) which is being given the opportunity to prove itself against the established winner. Note that while conceptually we have discussed two groups, there is no reason why these types of tests need be limited to only two tactics, and tests can easily be designed to simultaneously compare the impact of three, four, ten, or however many tactics there are.

Without this type of structured testing, you cannot observe the alternative states in which you are interested, even retrospectively! You will never actually know the answer to the question “should we have mailed green envelopes instead of red ones?” because if you only mailed the red ones, you won’t have any way of determining (beyond guesswork) what would have happened if you had mailed the green ones. This is the key point about testing---it allows you to quantify, with some degree of certainty, differences in outcomes arising from differences in your actions. Of course, many business people THINK they know what would have happened, but without any empirical basis it’s just a matter of one person’s hypothesis against another’s, and there is no real way to validate these judgments. In any large organization, there are bound to be differences of opinion regarding what course is going to lead to better outcomes. Testing provides a common way to resolve those differences of opinion by gathering relevant data and deciding on that basis. Without that data, other forms of resolution will certainly occur, because the organization will have to do something, but the basis of the decision will be on factors such as who has the most seniority, who argues the most persuasively or the longest, who has the most direct responsibility for the area in question, etc.. Unfortunately, none of these methods is guaranteed to produce the best outcomes for the organization, and testing sometimes reveals that what was believed to be the “best” treatment is in fact not optimal for the business.

This point is particularly important as the size and complexity of the decision space grows. When you are dealing with thousands of customers (or more), and evaluating the potential impact of millions of different decision points, it is highly unlikely that the same answer is going to be “best” in every instance. Likewise, as data analysis becomes more common in business and the availability of data improves exponentially, the ability to develop better solutions than a “one-size-fits-all” approach increases dramatically. It is rarely the case that the BAU strategy is actually the most optimal use of resources for each and every segment of the customer base. So testing allows for differentiation in treatment based on empirical knowledge of likely outcomes rather than *a priori* assumptions of how treatment should be varied between customer segments.

Another critical form of test to perform is usually referred to as the “holdout” test, which takes one group of targets or customers and keeps them out of whatever tactical plan you are designing. On its surface, this seems like a silly thing to do---after all, why would you not want to communicate with a group of potential targets or customers? But having a holdout group (often called the “control” group) is essential if you want to know how much your activities actually change your customer behavior, and the associated financial results for your business. Only testing can truly reveal the incremental impact of your activities compared to not doing anything at all. Any business decision-maker who is interested in questions of return on investment or the net present value of activities should be vitally concerned with this question. You don’t just want to know whether you should use the red envelopes versus the green envelopes---you want to know whether you should mail anything at all! Or more precisely, you want to know what is the marginal impact of sending mailings versus not sending mailings, so you can compare their impact to the cost of sending them in order to determine whether it is a worthwhile activity for your business.

Without holdout control groups, there are a number of ways that businesses attempt to measure the marginal impact of their activities. However, all have significant flaws in their design, because all are to some extent ignoring the basic logical fallacy with which we began our discussion of this objection. Probably the most common approach to estimating business impact without holdout groups is based on forecasting, and it runs something like this:

1. develop “baseline” forecast of expected results based on prior experience and BAU
2. implement a new or changed tactic in place of (or in addition to) BAU
3. measure any differences against the baseline forecast and attribute them to the change

Now, in a perfectly static world, this forecasting approach is not without some merit, and it is certainly better than ignoring the question of marginal impact altogether. However, it is important to identify the problems with this approach. First, it is only as accurate as the baseline forecast is accurate. But most businesses would have a difficult time achieving a high degree of forecast accuracy, even if they didn’t change anything in their business at all. So immediately, these businesses must recognize that some of what they are measuring as variance from their forecast is due to forecasting error and some of it is due to the different effect of the new tactic. Separating those effects and quantifying them for the purposes of an NPV analysis is not easily done and subject to the entire range of debate described previously, since ultimately there is little empirical basis for deciding that question.

Second, it really only allows the joint effect of multiple changes to be measured, rather than isolating the impact of any single change. Imagine a business that decides this year it is going to mail green envelopes instead of red ones, but it is also going to change the mailing frequency from 4 times per month to 3 times per month to reduce expenses, and it is also changing the price of the product being offered because of increased marketplace competition. This is not an unrealistic scenario, yet without a structured testing plan and holdout groups, there would be absolutely no way of attributing any observed effects specifically to any one of the three

changes that were made in their tactics. Once again, someone could come up with a framework for doing so, but it would be largely based on judgment and difficult to verify in any empirical way.

Another common approach to the question of identifying marginal impact is to do a comparative analysis between “responders” versus “non-responders” after a particular offer has been made. However, this type of analysis is entirely misleading, because in any set of customers there are already significant differences present before the business undertakes any activity at all, and the analysis of *ex post facto* group segments such as responders and non-responders may be driven by pre-existing differences rather than differences created by the activity in question. In other words, some people would have purchased your product, even if you didn’t send them any offers. But when you send an offer, and count all the activity of the responders as the outcome of the activity in question, or compare it to the activity of the non-responders and assuming the difference is entirely due to your activity, you are ignoring this fundamental point about pre-existing differences. So comparisons between responders and non-responders are highly misleading and almost always overstate the impact of the tactic being evaluated.

A third approach is to focus on things that can be measured and assume that they serve as good proxies or surrogates for the underlying change in behavior that is really the question. Using typical marketing metrics like response rate, open rate, conversion rate, click rate, etc., do not necessarily indicate the optimal strategy for the business, because higher values for these metrics do not tell you what would have happened if you did not pursue your BAU strategy. This is not an argument against collecting those metrics and using them to learn valuable information about how your marketing communications are being received, but it is a statement that such metrics can never really quantify the impact of the activities themselves compared to doing nothing at all, and that question is at the heart of evaluating the effectiveness of any business tactic.

Now that we have reviewed the most essential ideas for discussing the topic of testing, in the next installment in this series, we will cover common objections to testing on the basis of expense and uncertainty.