

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

**Part 3: How to Make Testing Successful**

Now that we have overcome the typical objections to testing (hopefully persuading you of the many important benefits of testing for improving your business decisions along the way), in this final installment of the series we will consider briefly the role of several key factors in the successful implementation of testing within any business.

Broadly speaking, the three stages of testing are as follows:

- A. Design
- B. Execution
- C. Analysis

We will not review each of these stages exhaustively, but rather we will focus only on specific elements that are particularly important to success and that experience suggests are those elements which are most frequently overlooked or implemented incorrectly. The steps outlined here are also compatible with more comprehensive approaches to analytics projects in general such as the CRISP-DM process, and many analytics teams will have a standard process flow. These are simply the most important conceptual elements that any such process flow should incorporate.

**A. Design**

In the test design phase, it is critical to begin by clearly defining the specific business question that you want answered before you set up the test. The more specific that this can be, the better. As already mentioned, the involvement of business analytics staff during the design phase is very important to ensure that the test is structured in a way that will enable the desired question to be answered with the greatest likelihood of success. Analysts will be able to help determine how large the test should be, how long it should be scheduled to run, how the samples should be drawn, and other technical points that are important to successful testing.

While operations personnel are not typically the best suited to designing tests, it is usually wise to include the operational staff that will be responsible for implementing a test before the design phase is finalized, because there may be operational considerations which would influence test design. It is a waste of time to design tests that cannot be implemented properly because of operational constraints, so including relevant resources in the design stage will minimize these occurrences. Part of the design phase of testing is the development of a detailed plan of action to define roles and responsibilities. Identifying someone in the organization who is going to take operational responsibility for making sure the test is implemented correctly is key. Analysts can design tests, but they are rarely the resources who are put in charge of executing the tests. If these two groups are out of alignment, then test failure is imminent.

Part of the design for any test should also be an estimation of the testing budget. As already described, the test budget should have two components: the direct and the indirect costs. Calculating how much the test is going to directly cost the organization in terms of marginal implementation costs is usually pretty straightforward (although remember that some testing may actually result in lower marginal costs than BAU). Added to these costs should be any incremental labor costs, setup costs, data costs, or other items that would fall outside the ordinary operating budget. However, it is important when analyzing the test results to be able to separate one-time costs associated with the test from ongoing costs associated with the corresponding tactic. For example, testing green envelopes when red envelopes are BAU may result in higher unit costs for green envelopes for the duration of the test because they are being produced in lower quantities. But if green envelopes were adopted as the BAU tactic and produced in higher volumes, then they would cost the same as the current red envelopes, so those cost differences would not be relevant to the comparison between the red and green envelope results. Only cost differences between treatments that would remain even if the test treatment were rolled out at full scale should be incorporated into the results analysis.

For indirect costs of testing, estimation is likely to be the only method available. Assumptions about the value of BAU tactics will need to be employed in order to quantify deviations from those tactics. For example, you might conclude (on the basis of prior testing) that the BAU tactic of mailing a catalog 4 times per month yields a revenue of \$12 per piece mailed, or \$48 per customer per month. You estimate in advance that the alternative scenario of mailing only 3 times per month will result in a value of \$15 per piece mailed, or \$45 per customer per month. Thus, the -\$3 difference in the monthly per customer revenue would be multiplied by the number of customers in the test times the number of months over which the test would run to estimate the opportunity cost of the test. Of course, the test group would likely have an associated printing, processing, and postage savings of more than \$3 per customer per month (the direct “cost” of the test) so the forecast of the test impact would be positive overall.

In many contexts, estimating the opportunity cost with any precision will be nearly impossible, so developing a probable range of outcomes and then conducting sensitivity analysis or Monte Carlo simulations to estimate the impact will help guide the budgeting process along. Ultimately, the goal is to develop an integrated test budget that incorporates both the direct and indirect costs of testing, and compares that against a projected benefit resulting from the alternative treatment being proposed in the test. In general, unless there is some plausible scenario through which the test would lead to an improvement in business results, by obtaining either better results at comparable costs, or comparable results at better costs, then there is no reason to pursue that particular test for the business.

## **B. Execution**

Be sure everyone involved in the execution of any test understands what is being tested and why. This is particularly important if part of the test involves a treatment which might be

considered “sub-optimal” by those involved, or if the test implementation is mediated through human contact (e.g., customer service representatives). You don’t want personnel making changes to treatments based on their own judgment of what is “best” and thereby invalidating the test results. Typically, once the relevant staff have been informed of the purpose of the test, they can be counted on to support it, especially if it is limited to a subset of customers and only runs for a specified duration.

The other essential point during execution of any test is that associated data capture is accurate and complete. If you design and execute tests flawlessly, but fail to collect the relevant data, then you will learn nothing at the conclusion! Part of the data collection work really must be done during the design phase to ensure that the important data elements are identified in advance and adequate provisions for their capture have been made. Then, early during the execution phase, check to make sure that the data are being collected in the manner expected.

Most good test designs encompass some degree of both active testing (structured testing) and passive testing (back-end learning) to enable the business to answer many different questions through the same test. An example of passive testing is the collection of data elements that are not directly related to the test design, but that allow the segmentation of results of the test on the back-end to determine whether the results differ across various segments of customers. With a large number of such potential elements, a single test could be used to look at many different possible segmentation schemes of the customer base. In some cases, the segmentation scheme is integral to the test design, and the groups are actually stratified in advance according to these attributes in order to attain size and representation goals. But in many other instances, the results will be divided up later based on the other data attributes that are available to determine whether there are any interesting patterns or results.

So the final recommendation in this area is that if you are in doubt, collect the data. In today’s business world, most data collection can be automated (rather than relying on manual processes) and data storage is cheap, so it is better to have too much data than too little data.

### **C. Analysis**

After a test is complete and all the data have been collected, it will be time for the analysts to take charge of the project again. However, be sure that the analysts doing the quantitative results analysis are very familiar with the original test design and purposes; ideally, they are the same people. There can be all sorts of idiosyncrasies in test design, such as the use of stratified sampling, and failure to take these factors into account can lead to results that are meaningless or (even worse) just plain wrong.

When analyzing and presenting test results, make sure the organization is focused on the right things. Analysts naturally want to spend lots of time reviewing the methodology and the intricacies of the technical work, but these are not very important to the business users. Assuming you have a capable analytics staff and adequate quality control processes in place,

these issues can be left in the background. You don't need to be a mechanic or understand all the details of automobile engineering to drive your car, you just need to trust that the person who put the car together was competent. Similarly, you don't need to be a statistician to look at test results, as long as you have one you trust who can vouch for the details behind the results.

Having said that, do let your statisticians or trained analysts help guide your interpretation of the results. Test results do not typically provide answers in black-and-white terms. They provide answers in terms of likelihood and probabilities---sometimes your test provides only weak evidence that one strategy is better than another, and your analytics staff will be able to help you understand what the results mean. Using test results to help make business decisions regularly will also cultivate the habit of thinking in terms of confidence intervals and uncertainty, which is probably helpful training in any decision-making context.

Finally, frame the test results in terms of what changes to the business would ensue if the test findings were applied. Quantify, whenever possible, but when doing so, be sure to incorporate persistent costs but ignore costs associated only with the limited implementation of the test itself (as described in the prior example of envelope cost differences).

Following these simple guidelines will help ensure your organization gets the most benefit out of testing and can use testing as a powerful tool to drive business success!