
1

Testing, Testing, Testing— Until You're Blue in the Face

Every once in a rare while, you'll hit the nail on the head with a powerful marketing concept—a seemingly perfect marriage of direct mail package and lists—at the very beginning of a new program. Your *control* package—the standard-bearing donor acquisition appeal you can keep mailing—may emerge more or less wholly formed from a successful pilot mailing.

I've seen that happen more than once.

But much more often it takes two or three tries, and months or years of step-by-step refinements, to produce a workhorse prospecting package that will build a large, responsive donor file for your organization. And then, sometimes quite suddenly, that package will “die,” and you'll have to come up with another one very quickly.

A donor acquisition program might not hit its stride until its third or fourth year or even later. The length of this process may

Testing, Testing, 1, 2, 3

relate to the organization's gradual accumulation of credibility and public recognition, to some shift in public sentiment about the issues the organization is addressing, or to changed attitudes in the organization's top management. But chances are it has a lot more to do with the cumulative value of thoughtful, systematic testing.

Now, here's what I mean by a *test*:

You might, for example, want to know whether the response to your prospect package will rise significantly if you suggest a lower minimum gift in the letter and on the response device. Currently, the package urges a contribution of \$25 or more; you decide to test whether suggesting a \$15 minimum will lift the response rate sufficiently to lower the overall donor acquisition cost (and without forcing the average gift to drop in proportion).

To determine this—reducing the question to its simplest form—you would take two equal and statistically equivalent groups of names from the same pool of prospect lists, mailing one group a package with the \$25 suggested minimum *ask* and the other with the \$15 *ask*.

The theory behind direct mail testing such as this is that applying scientific principles will enable you to discover the perfect combination of *ask*, offer, package, and postage—and ride off into the sunset, forever financially secure.

The reality's a little different.

For one thing the possible tests in a direct mail fundraising program are innumerable. It's sometimes excruciatingly difficult to figure out what's important to test and what isn't—particularly when your testing budget is strictly limited. In a program of modest scope it's especially important to test only those elements that are *significant*. Too much testing can eat up all the profits from your direct mail program.

Testing Until You're Blue in the Face

A wholly new acquisition package built around a new marketing concept is likely to be a significant test. Most of the time, so are meaningful variations in the benefits offered to donors. But even such variations might be insignificant under some circumstances.

Another problem with the proposition that rigorous testing will show the way to optimal results is that it's often difficult to design tests that are statistically meaningful. In most situations it requires many tens of thousands of names to establish confidence in the results of such things as price or postage tests. And even when you're able to test in these relatively large quantities, you'd be well advised to view the results with caution—no matter what the statisticians tell you.

Time and again I've seen results vary 20 percent or more between even very large samples when the variation was supposed to be no greater than 5 or 10 percent. According to experts in statistics, variations this great shouldn't occur more often than 1 to 5 percent of the time. I suspect they *often* occur, and so do some other direct mail fundraisers.

To test the validity of this proposition, on several occasions my colleagues and I arranged what we were assured were rigorous conditions: we split large groups of letters into two statistically identical groups to see what effect random variation might have. There was *no* difference between the two groups in package, postage, list, or any other controllable factor. All we were testing was the extent to which two equal prospect samples would produce equal results when simultaneously mailed identical packages. On both occasions, random variation had a greater effect than statistical theory said it should. Other consultants have had similar experiences.

No doubt there's some obscure theory of statistics that can explain away even these anomalous experiences. Rather than chase

Testing, Testing, 1, 2, 3

it down, however, I prefer to hold my clients' losses to a minimum and trust as little as possible in statistics—except when mailing quantities and testing budgets are big enough to add an extra margin of safety.

Why, then, when individual results may not be statistically reliable, should we bother to test such things as ask amounts, teasers, or postage? The answer is simple: by and large, *testing works*.

Testing works when we fundraisers test with *expectations* and then incorporate the results of our tests into future packages when our expectations are confirmed. Over time—sometimes after retesting vital elements two or three times—we're likely to learn valuable lessons that will enable us to upgrade a tolerably workable control package into one that is truly responsive.

Through testing—lifting response by 10 percent with this little change and 5 percent with that one—we may eventually cut the donor acquisition cost by 50 or 75 percent. Testing may, and often does, bring new life into a prospecting program by permitting us to expand the mailing volume and build large donor files relatively quickly. In large donor resolicitation programs, testing can also sometimes dramatically increase net revenue by establishing the most cost-effective use of postage, packaging techniques, suggested contributions, and the like. In this approach to direct mail prospecting, everything has the potential to be tested. As the following perspective section for this chapter suggests, even what everyone accepts as true may benefit from a reality check.

Testing is not limited to large direct mail programs. Almost any mailing of 10,000 names or more affords an opportunity to test. Smaller quantities limit the options and make testing a longer process. They demand extra care and greater patience, but the payoff can be just as dramatic in the end.

PERSPECTIVE: Shibboleths?

Conventional wisdom in direct mail fundraising maintains that a credit card option on the response device will depress results. It also says that a real stamp (a *live stamp*) on the package will boost response in comparison to the response to a package with a printed postal indicia or even a postage meter imprint.

'Tain't necessarily so. Not anymore, anyway. In a recent test for a client, my firm ran a head-to-head test of a check or credit card offer versus a check-only offer. There was no significant difference in response. Clearly, this is consistent with the experience of many major nonprofit mailers, who now routinely accept gifts by credit card. The same could be true for your organization—maybe.

In a test for another client, we compared a live, nonprofit bulk stamp with a postage meter imprint, mailing 30,000 packages in each fashion and expecting the stamp to outpull the meter, at least by a little. But the opposite proved to be the case. The meter outpulled the live stamp by 170 to 138 returns—a difference that was statistically significant. In other words these findings suggest it's not worth paying the extra cost of affixing a live stamp—at least it wasn't for that client at that time.

So, what conclusions can fundraisers draw from knowing that test results may not support what they have been assured is typical? Three conclusions, I believe:

1. Don't take conventional wisdom for granted.
2. Things may change over time.
3. Continual testing pays.

Testing, Testing, 1, 2, 3

And remember, a *test* is just as real a mailing as one that's called something else. One nonprofit's small test is another's huge prospect mailing. But both have to be evaluated in terms of costs and benefits.

Be sure, however, that you don't fall prey to the common but costly mistake of evaluating test results only in terms of their actual costs. Testing, as you will see in some of my examples, is often expensive. It may involve added creative, production, and even management costs—all of which need to be factored out when analyzing the outcome. Otherwise, you'll be comparing expensive apples to cheap oranges. Testing costs should be considered part of the overhead of your nonprofit's direct mail fundraising program as a whole. In fact, it's wise to add a line allocated specifically to testing to your direct mail budget.

Now, once you've designed and conducted your test, you've got to read and use the results, and that may be more easily said than done. The results from your test packages may conflict with the findings of earlier tests or with each other. The results will also be open to varying interpretations, depending on each interpreter's goals for the test. Furthermore, errors made in the *lettershop* (where your mailing is addressed and assembled) may call the validity of the results into question.

But the *biggest* danger in testing is that the uncertainty caused by equivocal results—or the anxiety raised by poor results—might tempt you to delay future mailings.

Important as your test results might be, the chances are that if you are trying acquire more donors for a productive, ongoing direct mail fundraising program, you'll be better off mailing blind than waiting for a definitive reading from tests—because your biggest mistake may be not to mail at all.