Final report
on SRG’s
Mobile Giving Project

Prepared by
Nick Allen, Mark Fuerst, and Marissa Goldsmith

December 2016
# Table of Contents

Introduction: Goals and Organization .................................................................................. 1  
Strategic Goals ......................................................................................................................... 1  
Project Organization ............................................................................................................... 1  
Operating Goals ..................................................................................................................... 2  
Project Activities ................................................................................................................... 3  
The Testing Process ............................................................................................................... 3  
  Initial Tests: removing non-essential fields ......................................................................... 4  
Round Two: Station-by-Station Recommendations ............................................................... 5  
General Test Findings ........................................................................................................... 6  
  1. Language Matters --but you have to remain focused on your ultimate goal ..................... 6  
    Case for WXPN .................................................................................................................. 6  
  2. Ask For Donations Where Users Are Entering Your Service ............................................. 7  
    Case for KJZZ ................................................................................................................... 7  
  3. Keep Donation Buttons Prominent on Mobile ................................................................. 7  
    Case for WXPN ................................................................................................................ 7  
  4. Remove Validation Steps (when possible) ......................................................................... 9  
    Case for WUNC .................................... ......................................................................... 9  
  5. Volume Matters ............................................................................................................... 9  
    Case for KING-FM ................................................ ..................................................... 10  
  6. Not All Best Practices Made a Difference ....................................................................... 12  
10 Applicable Results and "Just Do It" Recommendations ..................................................... 13  
Additional Observations, comparing participants to other NPOs ........................................... 15  
Major Challenges for Public Radio/TV .................................................................................... 17  
Appendix 1: Webinar Schedule .............................................................................................. 18  
Appendix 2: Final Tips for Setting up a Testing Practice ....................................................... 21  
Appendix 3: Testing Software ............................................................................................... 22  
Appendix 3: Sample Testing Recommendations .................................................................. 23  
  For WXPN .......................................................... .................................................................. 23  
  For WUNC .......................................................... .................................................................. 24  
  For KJZZ ................................................................. .................................................................. 25  
APPENDIX 4: Institutional impediments to testing ............................................................... 26
Introduction: Goals and Organization

In January 2016, the Station Resource Group (SRG) launched a new effort to advance SRG member-station and general system understanding of the emerging world of mobile fundraising.

The resulting project, later titled the **Mobile Giving Project**, had several strategic goals and three specific operating targets.

**Strategic Goals**

On the organizational (and "system") level, the Mobile Giving Project was aimed at filling a growing gap in fundraising R&D.

For more than a decade, CPB research investments had been aimed primarily at journalism and multi-station collaboration/consolidation. Relatively little investment was directed to fundraising research, especially for fundraising with a specific public radio focus. This was a marked change from the prior decade, when CPB funding helped to advance development efforts through a series of investments in

- Basic Membership Fundraising Practices (Brilliant on the Basics)
- Online and Email Marketing Practices (through investments at IMA)
- Major Gift Fundraising (Leaders Partnerships, Major Giving Initiative)

In the absence of CPB investment, advances in membership practice came primarily through the efforts of individual stations and leading consultants, including Lewis/Kennedy and Appleby/Arganbright, who championed the use of sustainer membership plans.

This left a large R&D gap in understanding mobile giving, a form of fundraising that rapidly advanced after 2007, initially through text-to-give donations to the Red Cross after the Haiti earthquake, then through to major mobile advocacy campaigns, and later through the Bernie Sanders and other political campaigns. All of this attention to mobile was accompanied by advances in responsive website design, making forms accessible on both desktop and phone-based browsers.

A few large stations--WNYC, KQED, WETA, WBEZ and others--were able hire senior staff with digital and mobile marketing experience. These teams advanced in-station R&D. But the vast majority of stations, including many SRG members, were not keeping pace.

**Project Organization**

Interest in a collaborative mobile-giving R&D effort surfaced during the 2015 SRG annual retreat. Work on the Mobile Giving Project started soon after that, beginning with an agreement to collaborate with the Public Media Futures Forum, which had devoted a large portion of its 2014 and 2015 agenda to analyzing changes in membership revenues, member file growth and member fundraising practices. Mark Fuerst, Director of the Forum project, agreed to assist SRG leadership organizing the new project, offering both staff time and some operating support provided by the Wyncote Foundation.
In mid-December 2015, Mark, Tom Thomas, and Terry Clifford selected Nick Allen, founder of Nuevo Fundraising and Open America, to be lead consultant for the project. Nick has an international reputation for digital fundraising with a focus on mobile giving. Prior to this project, Nick’s client list included NPR, UNICEF, Amnesty International, and The Humane Society of the United States.

Nick then selected Marissa Goldsmith to lead the day-to-day mobile testing. Marissa is a digital consultant, specializing in the use of data to improve online fundraising and marketing results. She has worked in web design and development in the nonprofit world for 20 years.

After several months of station recruitment, the Mobile Giving project hosted its first webinar on April 4, 2016, with nine participating stations:

- St. Louis Public Radio
- KING-FM, Seattle
- Wisconsin Public Radio
- WXPN-FM, Philadelphia
- New York Public Radio
- New Hampshire Public Radio
- WUNC-FM, Chapel Hill
- KJZZ and KBAQ, Phoenix
- WBEZ-FM, Chicago

The project team initially hoped to kick off the project with an in-person meeting, perhaps at an event or conference where many stations would already be represented. Given the project timing, that was not feasible. But we recommend an in-person kick-off meeting for similar projects as a way to share goals and difficulties, get face-to-face buy in, and build team spirit.

Operating Goals

From the start of the project there was a high level of agreement among participating stations about the importance of this research and the potential value it held. Everyone recognized that more and more listeners use their smartphones to read station emails, visit station web pages, and make donations. For some stations, 65% of their emails are opened on mobile phones.

There was also general agreement that the existing mobile forms, systems and marketing used by public radio needed to be advanced and improved.

This view was confirmed by our initial review conducted by Nick and Marissa of the mobile giving user experience. We found that a few stations had relatively advanced mobile design—with testing procedures in place— and all of the participating stations could benefit from an improved user experience.

The project team, in consultation with SRG, developed three primary objectives: Create more effective Smartphone donation pages that could

- Increase conversions (the percentage of people who come to the page who actually make a gift),
- Increase the number of sustainer joins over single-payment annual gifts, and
- Increase the average gift transaction.

Nick and Marissa then translated these goals into a set of project activities that guided most of the work. Working collaboratively with participating stations, the project would:
• Assemble and analyze baseline data about the volume of mobile donations and the online donation systems in place at these stations;
• Conduct a range of optimization tests on donation forms, both mobile Web and in-app;
• Investigate “text to pledge,” which encourages listeners to initially respond via text, and then receive a link that will take them to a donation page designed around the specific program content that has prompted them to want to give;
• Evaluate sustainer giving via text (PSMS) – the systems associated with making a donation or purchase by text message that is billed to the giver’s mobile phone bill;
• Review and possibly test payments via PayPal and Amazon;
• Develop a detailed overview of emerging best practices for mobile giving that can be referenced by stations’ staff moving forward.

Project Activities

The initial work scope called for 6 months of activity, starting in April and ending in September. Nick and Marissa conducted an initial station-by-station evaluation, to identify key testing opportunities. All stations provided access to their Google Analytics (GA), so that Marissa could evaluate current performance, as measured by GA tracking.

The majority of project work consisted of "A/B testing," using Optimizely software, with a focus on converting donors who visited donation pages on their smart phones. Our primary goal was to increase conversions, that is, increase the percentage of visitors who go to a donation page and actually complete a donation.

At Marissa’s recommendation, Optimizely was selected as the key testing platform. Nick contacted NPR Digital to discuss operational and organizational collaboration.

Bi-weekly project webinars were scheduled starting April 19. The majority of webinars focused on discussing proposed tests and reporting test results. Later, we invited guest speakers who could provide insight into the testing process and mobile giving experience, including

• **Nick Allen**: The Humane Society of the United States & Mobile
• **Ian Thiel, Optimizely**: Donation Optimization on Mobile
• **Marissa Goldsmith**: Campaign Tagging for Google Analytics
• **Dick McPherson**: How Email Strategy Affects Online Fundraising Results
• **Emily Patterson, PBS Digital**: Data and the User Experience

Agendas for the majority of webinars are contained in Appendix 1.

The Testing Process

At its core, A/B testing is involves the comparison of two alternatives (layouts, words, offers, choices) to learn which option works better. The alternatives can be as simple as two different words, such as "donate" or "pledge." Through A/B testing, we would try to determine which word produced more visits to a pledge form and led to more fully-completed donations. Many of the tests involved aspects of visual layout, in which case we would look at what happens if we remove one or more non-essential fields? Or what happens if we add a new field?
Ultimately, the goal is to increase in gross revenue. But we also wanted to encourage a "culture of testing," which, we found, is lacking in some parts of the public radio industry. We also encouraged stations to test items where the learnings could be shared with other stations.

For most of the station our tests went like this: The consulting team, with Marissa taking the lead in many cases,

- reviewed the existing mobile and online pledge forms and systems for each participant;
- developed a set of initial recommendations, based largely on the consultants’ experience from other mobile and online fundraising work;
- drew up a set of design/system recommendations for improvement; and
- established a set of testing recommendations, which then formed the basis for individual station A/B testing through Optimizely.

Some—but not all—stations tested these suggestions, and in the webinars we would analyze and report the results. The findings are discussed below. In some cases, stations acted quickly on the recommended changes, sometimes without testing. In those cases, we conducted a comparative “before and after” analysis.

**Initial Tests: removing non-essential fields**

Many of the participating stations had similar challenges with their donation forms, so, at first, we attempted to a "unified agenda" program, putting all participants on a similar testing schedule, focused primarily on removing non-essential fields from donation and pledge forms.

Implementing tests required stations to place Optimizely code on their web pages. Once the code was in place:

1) Marissa set up the test in Optimizely;
2) The stations were sent preview links to approve the test;
3) Upon approval, Marissa conducted a more thorough Quality Assurance (QA) process, using multiple browsers and devices, and completing transactions to ensure that they registered properly on the backend;
4) Once this QA was completed, Marissa worked with the station to set a “launch” date;
5) Once launched, Marissa would monitor the tests;
6) Once the tests were completed (either pass, fail, or after a set amount of time, no results), Marissa would write up the conclusions and recommended next steps.

Some tests were inconclusive, for two reasons: First, the measured change was not large enough to move the needle in a particular direction; second, the test shows a slight performance improvement, but the transaction volume levels were too low for statistically significant conclusions.

This initial round of tests demonstrated two key problems with a unified testing agenda. First, there were large differences in transaction volume. One station could complete a test in 4 days; another station with lower volume would need a month to accumulate the needed volume. Second, there were distinct differences in capacity and response. Some stations would quickly inset code and implement tests; others experienced obstacles and technical difficulties.
Round Two: Station-by-Station Recommendations

After the first round of testing, we abandoned the "unified" approach," and treated each station separately.

Marissa reviewed each station’s website and donation forms and identified 5 to 20 elements for testing. Using volume calculations, she provided each station with a prioritized testing list. The station would then select what to test. A sample, for WXPN, is shown below. Other samples of testing recommendations are contained in Appendix 2. This process led to a total of 10 tests conducted on the Optimizely platform. The tests fell into the following categories:

<table>
<thead>
<tr>
<th>Removal of non-essential, non-required fields</th>
<th>Test of Donation Language on Main Website</th>
<th>Reducing Steps in the Donation Form</th>
<th>Default to Credit Card instead of Automatic checking withdrawal</th>
</tr>
</thead>
<tbody>
<tr>
<td>WUNC, NHPR, WXPN, KWMU, KJZZ/KBAQ, KING</td>
<td>WXPN, KWMU</td>
<td>WUNC</td>
<td>NHPR</td>
</tr>
</tbody>
</table>

In addition to these tests, at least two stations, KJZZ/KBAQ and WXPN, made the changes that we had recommended without testing.

We planned three additional tests, but technical and timing issues prevented us from completing those efforts.

**Test Recommendations for WXPN**
General Test Findings

1. Language Matters -- But You Have To Focus On Your Ultimate Goals.

One of the easiest tests to conduct is testing the language (the words or phrases) that appears on a donate button. Do you use the word “Donate” or “Join” or “Pledge”? Which word generates more clicks and more gifts? We conducted this test for two stations, WXPN and KWMU. The results were surprisingly complex.

Case for WXPN

We conducted two rounds of testing for WXPN using all three options: "Pledge," "Donate," and "Join." The first round of testing, compared “Join” vs. "Donate." And in this initial round one, "Join" garnered the most clicks. In round two, we pitted “Join” against “Pledge.” In this round “Pledge” got the most clicks.  

Screen shots of the variations appear on the right.

This test series demonstrated the complexity of A/B testing when donation completion is the ultimate goal (vs. clicks to the donation page.)

But, regardless of which word generated more clicks to the pledge page, “Donate Now” provided the highest conversion rate.

The conversion rate for users who clicked either "Pledge" or "Join" and then completed their donation process was 20%. But the conversion rate for users who clicked "Donate" and completed the process was 35%. So while Pledge attracted the best level of click-through (vs. Donate or Join), Donate produced a best conversion rate.

We can only speculate about the reason for this: maybe users were not sure what “pledge” meant, and didn’t want to donate when they got to the donation form. Or donors may be responding to words they

---

1 We ran a similar test for KWMU that found no discernible difference in click-through rates for the three variations, perhaps because the volume of transactions was too low.
are most used to hearing on the air. Moreover, different words may work differently during different times of year. Donate almost always works better for year-end giving, or appealing to current sustainers.

With A/B testing often you don’t know why users did what they did, you just know they made some clicks at statistical significance!

2. Improved Results When Buttons are Positioned Where Users Are Entering

For many stations, news article pages and “Listen Live” pages are actually the home base of their users. Visitors often get to these pages without navigating from the home page. But sometimes these individual web pages contain no "ask." We hypothesized that adding a donate button to these pages would increase donations.

Case for KJZZ

KJZZ did not test this theory--they just implemented it, as shown on the right. Since this "Donate" rectangle button was included on the "Listen" page, 165 additional visitors went to the donation page. While we cannot attribute donations directly to clicks of this button, but if we assume the conversion rate is similar to all visitors (using the KJZZ conversion rate of 31%), then adding this button resulted in 51 additional donations.

3. Improved Performance when Buttons Are Properly Displayed in Mobile Layouts

A common pitfall in mobile design is to generate the mobile image using a sequence where all of the elements of the page appear in a sequence determined by the desktop-oriented layout. By using this approach, a donate button, which may be prominent and visible on a desktop, can almost disappear when the layout resizes for a mobile screen. This was an issue for WXPN.

Case for WXPN

The "Pledge Now" image, so clearly visible in the right side of the WXPN.org desktop presentation, disappeared when the same page was loaded on a Smartphone, as shown on the left.
To see the "Pledge Now" button a user would need to scroll down or open up the menu.

This problem was solved by placing an additional button in the mobile header, as shown below, right, with good results.

The chart below shows the test results measured by the rate of donation page sessions as a percentage of total sessions. Prior to inserting the new, more visible image, only 0.5% of mobile sessions led to a donation page visit. After the change, 1.2% of mobile sessions involved donation page visit. Because this change had no impact on the desktop site, we used the traffic to the desktop donation page as the control for our measurement. On desktop, visits increase by 109% during the test period (a pledge drive). In contrast, mobile visits increased by 153% and table visits increased by 186%.

Adjusting the mobile and table layout to display the donate button made a large measureable difference.
4. Improved Performance When Validation Steps Are Removed

Many tests in both retail and charitable contribution settings we find that the more steps required for completion of a purchase or contribution, the lower the completion rate. Our testing demonstrated this general rule.

Case for WUNC

Prior to our project test, WUNC used a three-step donation form:

Step 1: Fill out the Form  →  Step 2: Confirm Your Donation  →  Step 3: Complete Your Donation.

We decided to test what happened if you removed the middle step, "Confirm Your Donation."

During the time when the test was active, the form without the middle step had a conversion rate 1.3% greater (9.6%) than the conversation rate with the step (8.3%). Therefore, we declared removal of the conversion step to be "the winner," and WUNC updated all the main donation forms this way.

The results have been significant. Since we completely changed the WUNC form in mid-September, their conversion rate has been 13%. In the months prior to the change, this rate hovered around 6%. See the graph below.


To get conclusive results, you need sufficient volume of transactions to reach statistical validity. Many stations at least at this point do not generate enough mobile transaction volume to meet this threshold. As a result, some tests did not have enough traffic or conversions to reach statistically valid conclusions.
Case for KING-FM

To cite one example: we tested the main KING-FM donation form, removing unnecessary and potentially confusing fields such as “payment frequency” and a “start date” for sustainers.

The original form is shown below. A proposed variation is shown on the following page.

*KING FM Original Form*
A second form variation had twice as many completions as variation #1, but in both cases, the traffic and conversion levels were too low to reach statistically definitive conclusions about the impact of removing the fields.

Results are shown in the chart below.

The results at KING-FM were not unique. Half of all tests we ran either did not have enough volume to arrive at statistical significance, or had inconclusive results after a period of time.

That does not mean there weren’t learnings to be had. In addition to the regular tests, we also implemented click tracking. In the example for KING, we tracked how many users clicked on the Company Giving option in the original. The answer: no one. So while we may not have gotten conclusive results for the test, we came away with a few ideas to work on.
6. Not All Generally Applicable Best Practices Made A Difference

In general, retail and charitable donation testing has shown that a *simpler and shorter donation page produces better conversion rates*.

However, this was not always the case with the participating stations in this project.

We tested the removal of non-required fields for nearly all the stations in the project and found either:

a) We did not get statistical significance to prove the point (as in the KING example above);

b) We had sufficient the traffic, but we found no difference in response; or

c) For one station, the shorter form actually performed *slightly worse*.

While testing can give us the “what,” we can only hypothesize about the “why.” Our best guess is this: Visitors to public radio station donation forms, especially those in smaller markets, are highly qualified prospects. These potential donors listen to the station (often for dozens of hours per week). They have heard pledge announcements for years and may have donated in the past. For these highly motivated donors, optimization doesn’t make a big difference – they will complete the form even if it is “too long.”

Regardless, we remain convinced that optimization will improve conversation rates, especially if stations are looking to entice new donors and new listeners into their mix of support.
10 Applicable Results and "Just Do It" Recommendations

While each of these recommendations can be tested, our experience of testing by other stations and other non-profits and commercial organizations, gives us confidence that these recommendations will lead to improved mobile fundraising performance.

1. **Apply the test results regarding display of “donate” or “join” buttons.** Make sure your donate/join shows on the first screen of your mobile site. One of our test stations realized that the mobile site didn't show any call to action. They didn't bother to test – they just added “donate” at the top right of the screen and immediately got some donations!

   For WXPN, adding that little donate button resulted in **153% increase** in clicks to the mobile donation page.

2. **Adjust layouts so they are not just "mobile-friendly" – they should be mobile-optimized.** Optimize your mobile pages to make it easy to fill out your form on a small screen using your fingers. Form fields and buttons often need to be bigger. You may need to reduce the number of premiums offered to reduce screen clutter. Look for and eliminate non-essential fields, such as "How did you hear about us?" or "What’s your favorite program?" and "Comments."

3. **Where feasible pre-fill forms for supporters clicking though from an email, or coming from a page where they are logged in.** Pre-fill their donation form with the info you have – first and last names...  

4. **Remove confirmation/validation pages and captchas.** If you’re presenting a “confirmation page” after people have filled out the donation form -- take it out now. In our test with WUNC, as well as tests with dozens of other nonprofits, we’ve see an instant improvement when we remove that page. Convio introduced this confirmation page more than a decade ago, but most marketers abandoned it long ago because it hurts. Captchas are bad for usability, and even worse for accessibility. Many stations have been victim to fraud, and have added Captchas to counteract it. Before you do it, though, test it. Make sure that the cost of the fraud you’re preventing isn’t coming at the sacrifice of actual members and donors.

5. **Reduce your page load time** by reducing page size. Every study shows the faster load time improves conversion rates and your search engine rankings. It’s even more important on mobile devices, where network connection speeds are highly variable.

6. **Know where people are coming from.** Add Google Analytics UTM tagging to your donation form URL so you know where donors are coming from. Do this for your NPR One donations, your app donations, everywhere.
7. **Make it easy to choose either sustainer or one-time gifts.** Most stations want donors to become monthly sustainers and present that as the default. But do you make it easy to switch to one-off giving for those who prefer that? There are several forms that offer the options clearly that you might want to test (see screen shots).

8. **Reduce choices, especially on mobile.** We all want to give our donors options, but 10 donation levels may be pushing it. Test reducing the number of dollar options – it may increase total donations and average gift size.

9. **Consider eliminating premium gifts.** During non-pledge times, see what happens if you remove gift options from your form. If it doesn’t hurt the number of donations or reduce average gifts, not offering gifts can take some of the burden off of your internal team and save fulfillment costs.

   *While testing form fields, we tracked how many users choose a gift. Only 5-6% of visitors to the donation page choose a gift.*

10. **Test this: should we be using ACH as default?** We recognize the value of getting monthly sustainers on ACH, where the money is taken out of their bank accounts each month. ACH sustainers provide the highest lifetime value, because people don’t change banks very often and the transaction charges are lower than for credit card gifts. But most people don’t know their bank account numbers, let alone their “routing code.” Just imagine trying to find your checkbook while you’re making a transaction on your mobile phone. It’s obviously a serious physical challenge creating many barriers to form completion. What’s more, many millennials don’t even have check books. That said: we recommend that stations, collectively, should test whether ACH actually produces better long-term results over credit card payment on mobile devices.
Additional Observations, comparing participants to other NPOs

To add additional perspective, in early June Marissa compared participant station conversion rates with two somewhat comparable nonprofits. She found that "conversion rates for participating stations, are, overall, not actually that bad. If 25% is the sweet spot, most of the stations in that range... XPN and NYC are a bit lower... but they also pushing many more unqualified prospects to their donation forms."

She then asked: How could we explain this "not that bad" conversion rate with forms that seemed to need improvement? Her hypothesis was: public radio donors go to our forms with a strong intent to donate.

Equally revealing was a comparison of mobile conversion rates vs. desktop conversation rates for each participant, shown on the graph below right. In this review, Marissa dismissed the KWMU number as "deceiving, because the amount of traffic is so small." But she drew attention to the higher conversation rate for mobile at WNYC, suggesting they may have "a super-customized mobile experience that the desktop folks might appreciate."

The goal, Marissa suggested, would be "for conversion rates between desktop to show almost no difference, because you've removed all friction from the experience."
Where participating stations appeared to be "behind traditional nonprofits is getting users to their donation form." The graphs on this page show the percentage of sessions that involve a visit to a donation form.

Contrasting the NPO results with participating stations, Marissa noted that "both Nonprofit 1 & 2 have serious email efforts to get donors. That explains some of the difference. But not all. So where there really is room to improve is to get more eyeballs on donation forms. And for some (but not all) stations, there is a TON of room to improve on mobile."
Major Challenges for Public Radio/TV

In the course of our project, we noticed issues that are likely to be widely shared across the large and diverse set of non-commercial broadcasting stations, including:

- **The great variance in CRM and landing page back-ends** used by the stations. This variance makes it difficult to synchronize tests across groups of stations; instead, we had to treat each like a separate “client.”

- **Unwillingness to even test things that someone else found successful.** Example: While getting donors on an EFT/ACH system where the donation is deducted from the bank account increases a donor’s lifetime value, it’s typically much harder to get American donors to use their bank accounts over their credit cards. The issue was what appeared to be a negative attitude towards testing. We were able to get one station to test this (with no significant results).

- **A focus on internal department issues (vs. improving the user experience).** So, for example, some stations seemed more concerned about what data fields their database admin expected than the usability of the form (and potentially more donations per visitor)!

- **A wide range of Google Analytics setups and expertise.** We needed Google Analytics to get deeper into the analysis, but Google Analytics setups varied among stations, with many stations employing unsophisticated setups.

- **Outdated technology.** One example was the use of the obsolete BBNC software at KING-FM. With BBNC, it was impossible to track conversions in Optimizely, because the “thank you” page had the same URL as the home page. It was also extremely difficult to make a truly “mobile-friendly” form.

- **University-based stations using Sage Exchange** (or something similar). Stations using Sage Exchange credit card processing can’t streamline or improve forms (or really even track user experience) because this system takes you off site and interferes with event tracking.

- **General scheduling and responsiveness of the teams.** For one station, we never actually ran a test. Getting the station’s digital services team to do something as simple as put an Optimizely snippet on the page was often very time consuming.
# Appendix 1: Webinar Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Webinar Details</th>
</tr>
</thead>
</table>
| Apr 18, 2016 | Mobile Giving Project: Initial introduction, review of app observations  
- Goals of the project  
- Getting baseline data  
- Preliminary ideas on optimization testing  
- Using Google Analytics and Optimizely  
- Tools for sharing info and managing the project  
- Calendar  
- Regular call schedule to keep the project moving  
- What else? |
| May 02, 2016 | Mobile Giving Project check-in and update on status of first-round tests, set-up instructions and testing recommendations now in circulation. Marissa discussed testing designs.  
- Test designs went out to all the stations. (Still waiting confirmation from some.)  
- Marissa has the tests loaded in Optimizely. But they only get activated when code goes on the page.  
- We'll hear who has successfully put the code on their page. And other issues... |
| May 16, 2016 | Mobile Giving Project continued review of first-round tests (some already in the field), set-up issues (if any) and testing recommendations already in circulation. Marissa reviewed Google Analytics tagging issues and status of tests.  
**NHPR** - In this test, we reduced the number of contact fields - primarily things like "title", "middle name", but also a slew of "join donor" fields. Although not conclusive, the test looked like it was going on to win, when suddenly, the donation form was completely changed out (to not include any of those fields).  
**WXPN** - In this test, we reduced some fields, but also hid fields if a user didn't want a gift. This test launched this week, amid the pledge drive. There is no conclusion yet, and so far the tests are event. But that is to be expected, because I made the split 80/20 since we launched in the middle of pledge. Another thing I didn't consider, is that most people who pledge during a pledge drive want a gift. So they are less likely to click the "No gift for me, thanks." button, which hid most of those fields. Also, they actually had a donation form outage (not related to us) on their last day of pledge.  
**KJZZ** - We reduced the number of fields. So far, this test is inconclusive.  
**KBAQ** - The test is more or less the same as KJZZ. Marissa had not properly targeted the URL, and so for the first day or two, tests |
were always being delivered, but they are now. The following tests are ready to go and are out for approval:

**KWMU**  
**KING**  
Code still needed on the page for:  
**WPR**  
**WUNC**  
And **WNYC** and **WQXR** will run the tests themselves.

**Jun 06, 2016**  
Mobile Testing Project check-in and ... Nick presented testing and mobile examples from his client Humane Society of the United States  
Marissa presented station tests:  

**NHPR.** These were the folks who took down their donation form without telling anyone. Yesterday, I was ready to start a new test for them, when I noticed the old form was back up. So I turned the test back on. Apparently, the new form had credit card processing issues. The new form was back up this morning, so I turned the test off. One extra day of testing put it closer to winning, but we'll never really know.

**KING.** This test launched this week. Right now, the test is winning, but their conversion rate is shockingly low on both. I'm going to get in touch with them to make sure what I'm seeing seems right.

**KBACH/KJZZ.** On KJZZ, the test is actually losing, but there's not enough traffic to say there's statistical significance. If it remains at the current pace, it will lose in 5 days. KBACH's traffic is so low, that nothing is registering. I am going to be in touch with them about this.

**KWMU.** This test was running for 2 days, when we had to pause it, because there was one form where it was hiding the wrong field. I've fixed it and am awaiting approval to re-run the test. For the two days it was running, the test was winning significantly, and, if it remained at the same pace, will win in 2 days.

**WXPN.** This is the one that's making me sad, because it's a huge market and I figured this would be done quickly. The original is only slightly beating the test, but really, it's a dead heat. If the trend remains the same, the original will beat the test after 23 days. That's longer than I want to run this test. So if there isn't any real change by the end of next week, I'm going to call it a draw.

**WPR.** The development team at WPR didn't agree with my ideas for the test. So I've proposed a general "tightening up" of the donation form. They seem more accepting of that. I just need approval to move forward.

**WUNC.** My contact was out of town last week, but promises to get the code on the page next week.
**WNYC/WQXR.** They are going to run their own donation string test starting this weekend, when pledge starts. Over the weekend, the pledge announcements will not push a particular donation string level, so we’ll get a good sense of what is working. Starting Monday, though, they will, so it will be a little less valid.

**CPM.** I'm awaiting your instructions on what to do next with them.

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun 20, 2016</td>
<td>SRG Mobile Giving Project check-in a...</td>
</tr>
<tr>
<td>Jul 11, 2016</td>
<td>SFG Mobile Giving Project check-in</td>
</tr>
<tr>
<td></td>
<td>• Brief update</td>
</tr>
<tr>
<td></td>
<td>Marissa Goldsmith</td>
</tr>
<tr>
<td></td>
<td>Donation Optimization Discussion</td>
</tr>
<tr>
<td></td>
<td>Ian Thiel, Head of New Products at Optimizely</td>
</tr>
<tr>
<td></td>
<td>• Look ahead to SRG Retreat and PMDMC</td>
</tr>
<tr>
<td></td>
<td>Mark Fuerst</td>
</tr>
<tr>
<td>Aug 22, 2016</td>
<td>SRG Mobile Giving Project update check-in</td>
</tr>
<tr>
<td>Aug 23, 2016</td>
<td>Dick McPherson presented &quot;Digital Fundraising Roadmap&quot; from PMDMC</td>
</tr>
<tr>
<td>Sep 12, 2016</td>
<td>SRG Mobile Giving Project update check-in</td>
</tr>
<tr>
<td>Oct 17, 2016</td>
<td>SRG Mobile Giving Webinar</td>
</tr>
</tbody>
</table>
Appendix 2: Final Tips for Setting up a Testing Practice

- **Keep an Ideas Pool.** Whenever you have an idea for a test, write it down. We recommend keeping a shared spreadsheet or Google Doc of ideas so that everyone can contribute. You can use the testing ideas spreadsheet we used in this project as a base. (It’s attached.)

- **Set a Calendar.** Use a calendar so you can set approximate dates for your tests. Use the Testing Priorities Spreadsheet you got from this project (attached) as a base to help you figure out how long tests should take.

- **Don’t go on forever.** Use the testing priorities spreadsheet to figure out how long a test should take. Don’t go too far beyond it. And don’t go on longer than a month. A test that is taking longer than a month may fall victim to seasonality. Turn it off and test something else.

- **Set a testing budget.** It’s likely that you will need to pay for testing software. Set your budget ahead of time. Since most testing software sets its pricing based on visitors, this budget will help you determine what you should test and how often.

- **Make sure your Google Analytics is (more or less) accurate.** Since your testing budget will be based on visitors, you want to make sure your Google Analytics is accurate. Make sure Google Analytics code is on all the pages you want to test.

- **Don’t run more than one test at a time.** Even the most sophisticated testing teams can run into trouble when running more than one test at a time. Keep it simple and only run one test at a time.

- **Pay attention to the “IF” phrases.** When collaborating, pay attention to phrases “I wonder if...” or “What would happen if...” These thoughts are testing fodder.

- **Identify your testing champions and cheerleaders.** You are part of a team. So get your team involved in testing. Ask for ideas, and show them the results.
Appendix 3: Testing Software

To properly conduct A/B and other testing, you will need testing software. Below are some of the more popular pieces of software available.

When evaluating software, there are a few things to consider:

- Some software offers **Landing Page** testing only. This means that you test one page against another. This is effective if you want to test a single donation page. It is less effective if you wanted to test something global, like the donate buttons across your website.

- Most testing software requires you to put a snippet of code on your website. Each may have its own set of rules. Before buying anything, make sure your IT team will allow you to implement the code.

- Some testing software will either charge you more for, or may not be able to handle, A/B testing and tracking *across domains*. For example, if your website is [http://www.wxpn.org](http://www.wxpn.org) and you want to test a donate button, but your donation form lives at [https://donate.nprstations.org/wxpn/](https://donate.nprstations.org/wxpn/), the tool may not be able to track a conversion from beginning to end. There are workarounds that usually involve Google Analytics and extra analysis. If you want to do cross-domain testing, make sure your software can handle it, or be prepared to spend a little more time in setup and/or analysis.

- Most testing software pricing is based on visitors per month. Think about what you want to test, and then look in your analytics to see how much traffic you get.

Here are some of the more popular software options. These are our opinions based on experience.

- **Optimizely.** Optimizely is one of the more popular testing tools. It is affordable for landing page testing, less so for global website testing. Their pricing model is ever-changing, but nonprofits can usually negotiate discounts. It is an extremely user-friendly tool. Tests can be extraordinarily simple. Or, with a little bit of jQuery knowledge, you can run very complex tests.

- **Visual Website Optimizer (VWO).** VWO is Optimizely's closest competitor in usability and price. It is a little bit easier to use for Cross Domain testing.

- **Unbounce.** Unbounce is strictly for Landing Page Testing with a focus on lead conversion. It shines in ease of use, and comes with many built-in landing page templates.

- **Adobe Target.** Adobe offers sophisticated enterprise testing. It is the most expensive of all the options, and works best if you use other Adobe Marketing Cloud tools.

- **Google Optimize 360.** This is Google’s new and improved testing platform. It is still in beta. The free version has limitations over the enterprise paid version.

If you don’t have a budget for software, but use Google Analytics with Google Tag Manager (GTM) and have some in-house tech skills, you can do some DIY testing. This [blog post by Simo Avaha](http://example.com) shows you how.
## Appendix 3: Sample Testing Recommendations

For WXPN

<table>
<thead>
<tr>
<th>Experiment Priority</th>
<th>Experiment Name</th>
<th>Primary Goal</th>
<th>Baseline Conversion Rate %</th>
<th>Target Conversion Rate %</th>
<th>Total Variations</th>
<th>Start Date</th>
<th>End Date</th>
<th>Daily Unique Visitors</th>
<th>Traffic Allocation</th>
<th>Planned Sample Size</th>
<th>Minimum Sample Size</th>
<th>Planned Days</th>
<th>Minimum Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reduce Total Number of Giving Strings</td>
<td>Increase average gift value.</td>
<td>15.00%</td>
<td>20.00%</td>
<td>2</td>
<td>1-Aug-16</td>
<td>15-Aug-16</td>
<td>408</td>
<td>100%</td>
<td>5,712</td>
<td>1,316</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Test the team Pledge vs. Donate</td>
<td>Increase total visits to pledge page.</td>
<td>4.10%</td>
<td>4.51%</td>
<td>3</td>
<td>1-Aug-16</td>
<td>15-Aug-16</td>
<td>7,375</td>
<td>100%</td>
<td>103,250</td>
<td>88,158</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Move Search Button Down and replace with Donate button.</td>
<td>Increase total visits to pledge page.</td>
<td>4.10%</td>
<td>4.51%</td>
<td>2</td>
<td>16-Aug-16</td>
<td>30-Aug-16</td>
<td>7,375</td>
<td>100%</td>
<td>103,250</td>
<td>58,772</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Add a Static Donation Call on Mobile &quot;Article&quot; Pages</td>
<td>Increase total visits to pledge page on mobile devices.</td>
<td>3.54%</td>
<td>3.99%</td>
<td>2</td>
<td>31-Aug-16</td>
<td>14-Sep-16</td>
<td>944</td>
<td>100%</td>
<td>13,216</td>
<td>68,390</td>
<td>14</td>
<td>73</td>
</tr>
<tr>
<td>Experiment Priority</td>
<td>Experiment Name</td>
<td>Primary Goal</td>
<td>Baseline Conversion Rate %</td>
<td>Target Conversion Rate %</td>
<td>Total Variations</td>
<td>Start Date</td>
<td>End Date</td>
<td>Daily Unique Visitors</td>
<td>Traffic Allocation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------------------------------------------</td>
<td>-------------------------------------------------------</td>
<td>---------------------------</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>------------</td>
<td>-----------------------</td>
<td>--------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Hide some giving strings on mobile devices</td>
<td>Increased transactions on mobile devices</td>
<td>1.07%</td>
<td>1.34%</td>
<td>2</td>
<td>2-Oct-16</td>
<td>1-Nov-16</td>
<td>141</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Make Donation 1-step</td>
<td>Increased transactions</td>
<td>7.87%</td>
<td>9.84%</td>
<td>2</td>
<td>1-Aug-16</td>
<td>31-Aug-16</td>
<td>183</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Remove Captcha</td>
<td>Increased transactions</td>
<td>7.87%</td>
<td>9.84%</td>
<td>2</td>
<td>1-Sep-16</td>
<td>1-Oct-16</td>
<td>183</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Make Custom NPR One Donation Form at lower end</td>
<td>Increased transactions from NPR 1</td>
<td>1.07%</td>
<td>1.34%</td>
<td>2</td>
<td>2-Nov-16</td>
<td>2-Dec-16</td>
<td>35</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sample Size Projection**

<table>
<thead>
<tr>
<th></th>
<th>Planned Sample Size</th>
<th>Minimum Sample Size</th>
<th>Planned Days</th>
<th>Minimum Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Hide some giving strings on mobile devices</td>
<td>4,230</td>
<td>38,082</td>
<td>30</td>
<td>271</td>
</tr>
<tr>
<td>2 Make Donation 1-step</td>
<td>5,490</td>
<td>4,806</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>3 Remove Captcha</td>
<td>5,490</td>
<td>4,806</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>4 Make Custom NPR One Donation Form at lower end</td>
<td>1,050</td>
<td>38,082</td>
<td>30</td>
<td>1,089</td>
</tr>
<tr>
<td>Experiment Priority</td>
<td>Experiment Name</td>
<td>Primary Goal</td>
<td>Baseline Conversion Rate %</td>
<td>Target Conversion Rate %</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------</td>
<td>--------------------------------------------------</td>
<td>----------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Test Donate Language</td>
<td>Get more visitors to the donation page.</td>
<td>0.62%</td>
<td>0.68%</td>
</tr>
<tr>
<td>2</td>
<td>Add Donate to Main Menu</td>
<td>Get more visitors to the donation page.</td>
<td>0.62%</td>
<td>0.68%</td>
</tr>
<tr>
<td>3</td>
<td>Reduce mobile giving strings</td>
<td>Increase total mobile transactions.</td>
<td>11.12%</td>
<td>13.11%</td>
</tr>
<tr>
<td>4</td>
<td>Add Donate to Listen Live</td>
<td>Get more visitors to the donation page.</td>
<td>0.62%</td>
<td>0.68%</td>
</tr>
</tbody>
</table>
APPENDIX 4: Institutional impediments to testing

by Nick Allen and Marissa Goldsmith

We encountered institutional impediments to A/B testing in this project. In fact, we have seen many of the same impediments at other nonprofit organizations we have worked with – and at plenty of for-profits and government agencies.

- **Siloed departments**: Digital and membership departments don't always know what each other is doing; in some cases, we are working with the digital department, and they are constantly going back and forth between us and membership (though once membership is involved, they are interested in the project).

- **The big stations want to do their own testing**: The largest stations have the staff and technology resources, which gives them a distinct advantage—and segments the group process. For example, WNYC staff tuned into the Web meetings, but didn't participate any of the project’s testing. We asked them to present findings of their own testing, but they were not able to schedule time. The small stations often don't have the staff, the expertise, the technology, and/or the budget.

- **Possible gap between the goals of senior leadership and the work assignments**: There seemed to be a divergence of attention(and often overload) between upper and lower-level staff that we are often working with.

- **Stretched-thin teams**: We talked about this, but the people we are dealing with on a near-daily basis have a lot going on. First, they say they aren't able to do much because they have a pledge drive. Then it’s year end fundraising. Then there's an event. And then the next pledge drive.

- **Digital bottlenecks**: Sometimes, to make a change, even a small one, stations have to jump through hoops and wait a long time. In 2016, a good CRM should enable authorized users to make changes in real time. But it’s not always like this for the stations we are working with. Sometimes, to add a tiny Optimizely script to an NPR Digital Services page, the station has to file a support ticket—and be told it can’t be done because of another project. Sometimes, a vendor may be required to make the changes. Other times more people than necessary have approve relatively small changes at the stations.

- **Payment processors outside the CRM system**: This is more a technical issue, mainly for stations that have university affiliations. We want to optimize the donor experience, but if the stations are required to redirect donors to an exterior donation page that usually is not optimized, we can only do so much. This is particularly a problem for those on the Allegiance platform.
• **Out-of-date technologies:** This is not necessarily an institutional impediment, but we would consider both Allegiance and Blackbaud’s Net Community tool to be out-of-date, but institutionally it might be a tough haul to get the stations to move onto another platform.

• **Lack of a culture of testing:** Of course, many organizations don’t have a developed culture of testing, but stations should prioritize development of a culture of testing because with digital and mobile, testing goes right to the bottom line. Fast, cheap A/B digital testing can improve a station’s marketing ROI by 10, 20, even 30% in a few months, and it’s going to be ever simpler to set up the tests and ever more important for acquiring online donors. For stations who are undergoing rapid changes in channels and content creation and delivery, testing is critical and it needs to be a regular part of daily operations, with enough trained staff and budget.
Supplement 1: Resources for Testing

- **Optimizely 2016 Testing Toolkit.** This toolkit contains spreadsheets, presentation templates, and other tools to move testing forward. Many of these should look familiar to you, as they were used throughout the course of this project.

- **ConversionXL: A/B Testing Mastery: From Beginner to Pro in a Blog Post.** This is a comprehensive guide to everything A/B testing.

- **GetDriven Statistical Significance Calculator.** This is one of many Statistical Significance calculators you can find online.

- **Optimizely Sample Size Calculator.** Tells you how big your sample size needs to be to get results.

- **Evan’s Awesome A/B Tools.** More statistical calculators to help you with your A/B testing.

Good Testing Blogs and Resources

There are great resources available if you are interested in keeping up-to-date with the latest in testing. Here are a few of our favorite blogs and authors:

- **Optimizely Blog:** [https://blog.optimizely.com/](https://blog.optimizely.com/).
- **VWO Blog on Conversion Rate Optimization:** [https://vwo.com/blog/](https://vwo.com/blog/).
- **Behave.org (formerly Which Test Won):** [https://www.behave.org/](https://www.behave.org/).
- **Unbounce Blog:** [http://unbounce.com/blog/](http://unbounce.com/blog/).
- **ConversionXL:** [http://conversionxl.com/blog/](http://conversionxl.com/blog/).

Resources for Mobile

- **Email on Acid:** [https://www.emailonacid.com/](https://www.emailonacid.com/). This program can test your emails in hundreds of OS/Email Client combinations. This can help you ensure that your emails look good in everyone’s inbox. Pricing starts at $45 per month. *Note:* If you use Blackbaud Luminate Online, you may have access to a limited number of Email on Acid tests at no charge.

- **CrossBrowserTesting.com:** [https://crossbrowsertesting.com/](https://crossbrowsertesting.com/). This program gives you remote access to thousands of different OS/Browser/Resolution combinations, so that you can test parts of your website to make sure they operate correctly on every device. You can do live tests, or do mass screenshots.

- **SitePoint Articles on Mobile UX:** [https://www.sitepoint.com/mobile/ux-mobile/](https://www.sitepoint.com/mobile/ux-mobile/). SitePoint is the premier blog for the tech savvy. They have a lot of great articles on Mobile UX.

- **User Interface Engineering (UIE):** [https://www.uie.com/](https://www.uie.com/). This is the company of UX Guru, Jared Spool. The articles and blog posts run the range of all devices, but the articles on mobile usability and responsive design are some of the best in the industry.
• Google’s Getting Started Guide to Mobile Friendly Websites: https://developers.google.com/webmasters/mobile-sites/get-started. This is outline by Google gives you some excellent guidelines for mobile experience.

• 7 Steps to Ensure Your Nonprofit Has a Mobile-Friendly Website: https://www.causevox.com/blog/mobile-friendly-nonprofit-website/. This blog post by Tina Jepson of CauseVox gives some guidelines for mobile-friendliness that are specific to the non-profit world.
SUPPLEMENT 2: The A/B Testing Checklist You'll Want to Bookmark

Written by Lindsay Kolowich

https://blog.hubspot.com/marketing/a-b-test-checklist#sm.000007n1mao8mud5ixlhvlx99hgyo

When marketers like us create landing pages, write email copy, or design call-to-action buttons, it can be tempting to use our intuition to predict what will make people click and convert.

But basing marketing decisions off of a “feeling” can be pretty detrimental to results. Rather than relying on guesses or assumptions to make these decisions, you’re much better off running conversion rate optimization (CRO) tests.

CRO testing can be valuable because different audiences behave, well, differently. Something that works for one company may not necessarily work for another. In fact, CRO experts hate the term “best practices” because it may not actually be the best practice for you.

But these tests can also be complex. If you’re not careful, you could make incorrect assumptions about what people like and what makes them click -- decisions that could easily misinform other parts of your strategy.

One of the easier (and most common) types of CRO tests is called an A/B test. An A/B test simply tests one variable in a piece of marketing content against another, like a green call-to-action button versus a red one, to see which performs better.

So, what does it take to run an A/B test, exactly? Keep reading to learn what an A/B test is in a little more detail, followed by a full checklist for what marketers should do before, during, and after these tests. You’ll want to bookmark this for your next one.
How A/B Tests Work

To run an A/B test, you need to create two different versions of one piece of content with changes to a single variable. Then, you'll show these two versions to two similarly sized audiences, and analyze which one performed better.

For example, let's say you want to see if moving a certain call-to-action button to the top of your homepage instead of keeping it in the sidebar will improve its conversion rate.

To A/B test this change, you'd create another, alternative web page that reflected that CTA placement change. The existing design -- or the "control" -- is Version A. Version B is the "challenger."

![A/B Testing](image)

*Image Credit: ConversionXL*

Then, you'd test these two versions by showing each of them to a predetermined percentage of site visitors. (To learn more about A/B testing, download our free introductory guide here.)

Now, let's walk through the checklist for setting up, running, and measuring an A/B test.

**Checklist for Running an A/B Test**

**Stage 1: Before the A/B Test**

1) Pick one variable to test.

As you optimize your web pages and emails, you might find there are a number of variables you want to test. But to evaluate how effective a change is, you'll want to isolate one, single variable and measure its performance -- otherwise, you can't be sure which one was responsible for changes in performance. You can test more than one variable for a single web page or email -- just be sure you're testing them one at a time.

Look at the various elements in your marketing resources and their possible alternatives for design, wording, and layout. Other things you might test include email subject lines, sender names, and different ways to personalize your emails.
Keep in mind that even simple changes, like changing the image in your email or the words on your call-to-action button, can drive big improvements. In fact, these sorts of changes are usually easier to measure than the bigger ones.

**Note:** There are some times when it makes more sense to test multiple variables rather than a single variable. This is a process called multivariate testing. If you're wondering whether you should run an A/B test versus a multivariate test, [here's a helpful article from Optimizely](https://www.optimizely.com) that compares the two.

2) **Choose your goal.**

Although you'll measure a number of metrics for every one test, choose a primary metric to focus on -- before you run the test. In fact, do it before you even set up the second variation. If you wait until afterward to think about which metrics are important to you, what your goals are, and how the changes you're proposing might affect user behavior, then you might not set up the test in the most effective way.

3) **Set up your "control" and your "challenger."**

Set up your unaltered version of whatever you're testing as your "control." If you're testing a web page, this is the unaltered web page as it exists already. If you're testing a landing page, this would be the landing page design and copy you would normally use.

From there, build a variation, or a "challenger" -- the website, landing page, or email you'll test against your control. For example, if you're wondering whether [including a testimonial on a landing page](https://www.optimizely.com) would make a difference, set up your control page with no testimonials. Then, create your variation with a testimonial.

4) **Split your sample groups equally and randomly.**

For tests where you have more control over the audience -- like with emails -- you need to test with two or more audiences that are equal in order to have conclusive results.

How you do this will vary depending on the A/B testing tool you use. If you're a [HubSpot Enterprise customer conducting an A/B test on an email](https://www.optimizely.com), for example, HubSpot will automatically split traffic to your variations so that each variation gets a random sampling of visitors.

5) **Determine your sample size (if applicable).**

How you determine your sample size will also vary depending on your A/B testing tool, as well as the type of A/B test you're running.
If you're A/B testing an email, you'll probably want to send an A/B test to a smaller portion of your list to get statistically significant results. Eventually, you'll pick a winner and send the winning variation on to the rest of the list. (Read this blog post for a more detailed guide on calculating an email A/B test's sample size.)

If you're a HubSpot Enterprise customer, you'll have some help determining the size of your sample group using a slider. It'll let you do a 50/50 A/B test of any sample size -- although all other sample splits require a list of at least 1,000 recipients.

If you're testing something that doesn't have a finite audience, like a web page, then how long you keep your test running will directly affect your sample size. You'll need to let your test run long enough to obtain a substantial number of views, otherwise it'll be hard to tell whether there was a statistically significant difference between the two variations.

6) Decide how significant your results need to be.

Once you've picked your goal metric, think about how significant your results need to be to justify choosing one variation over another. Statistical significance is a super important part of A/B testing process that's often misunderstood. If you need a refresher on statistical significance from a marketing standpoint, I recommend reading this blog post.

The higher the percentage of your confidence level, the more sure you can be about your results. In most cases, you'll want a confidence level of 95% minimum -- preferably even 98% -- especially if it was a time-intensive experiment to set up. However, sometimes it might make sense to use a lower confidence rate if you don't need the test to be as stringent.

Matt Rheault, a senior software engineer at HubSpot, likes to think of statistical significance like placing a bet. What odds are you comfortable placing a bet on? Saying "I'm 80% sure this is the right design and I'm willing to bet everything on it" is similar to running an A/B test to 80% significance and then declaring a winner.

Rheault also says you'll likely want a higher confidence threshold when testing for something that only slightly improves conversation rate. Why? Because random variance is more likely to play a bigger role.

"An example where we could feel safer lowering our confidence threshold is an experiment that will likely improve conversion rate by 10% or more, such as a redesigned hero section," he explained. "The takeaway here is that the more radical the change, the less scientific we need to be process-wise. The more uber-specific the change (button color, micro copy, etc.), the more scientific we should be because the change is less likely to have a large and noticeable impact on conversion rate."

7) Make sure you're only running one test at a time on any campaign.

Testing more than one thing for a single campaign -- even if it's not on the same exact asset -- can do a number on your results. For example, if you A/B test an email campaign that directs to a landing page at the same time that you're A/B testing that landing page ... how can you know which change caused the increase in leads?
Stage 2: During the A/B Test

8) Use an A/B testing tool.

To run an A/B test on your website or in an email, you'll need to use an A/B testing tool. If you're a HubSpot Enterprise customer, the HubSpot software has features that let you A/B test emails (learn how here), calls-to-action (learn how here), and landing pages (learn how here).

For non-HubSpot Enterprise customers, other options include Google Analytics' Experiments, which lets you A/B test up to 10 full versions of a single web page and compare their performance using a random sample of users.

9) Test both variations simultaneously.

Timing plays a significant role in your marketing campaign’s results, whether it's time of day, day of the week, or month of the year. If you were to run Version A during one month and Version B a month later, how would you know whether the performance change was caused by the different design or the different month?

When you run A/B tests, you'll need to run the two variations at the same time, otherwise you may be left second-guessing your results.

The only exception here is if you're testing timing itself, like finding the optimal times for sending out emails. This is a great thing to test because depending on what your business offers and who your subscribers are, the optimal time for subscriber engagement can vary significantly by industry and target market.

10) Run the test long enough to get substantial results.

Again, you'll want to make sure that you let your test run long enough in order to obtain a substantial sample size. Otherwise, it'll be hard to tell whether there was a statistically significant difference between the two variations.

How long is long enough? Depending on your company and how you execute the A/B test, getting statistically significant results could happen in hours ... or days ... or weeks. A big part of how long it takes to get statistically significant results is how much traffic you get -- so if your business doesn't get a lot of traffic to your website, then it'll take much longer for you to run an A/B test. In theory, you shouldn't restrict the time in which you're gathering results. (Read this blog post to learn more about sample size and timing.)

11) Ask for feedback from real users.

A/B testing has a lot to do with quantitative data ... but that won't necessarily help you understand why people take certain actions over others. While you're running your A/B test, why not collect qualitative feedback from real users?

One of the best ways to ask people for their opinions is through a survey or poll. You might add an exit survey on your site that asks visitors why they didn't click on a certain CTA, or one on your thank-you pages that asks visitors why they clicked a button or filled out a form.
You might find, for example, that a lot of people clicked on a call-to-action leading them to an ebook, but once they saw the price, they didn't convert. That kind of information will give you a lot of insight into why your users are behaving in certain ways.

**Stage 3: After the A/B Test**

12) **Focus on your goal metric.**

Again, although you'll be measuring multiple metrics, keep your focus on that primary goal metric when you do your analysis.

For example, if you tested two variations of an email and chose leads as your primary metric, don't get caught up on open rate or clickthrough rate. You might see a high click through rate and poor conversion rates, in which case you might end up choosing the variation that had a lower click through rate in the end.

13) **Measure the significance of your results using our A/B testing calculator.**

Now that you've determined which variation performs the best, it's time to determine whether or not your results statistically significant. In other words, are they enough to justify a change?

To find out, you'll need to conduct a test of statistical significance. You could do that manually ... or you could just plug in the results from your experiment to our free A/B testing calculator. For each variation you tested, you'll be prompted to input the total number of tries, like emails sent or impressions seen. Then, enter the number of goals it completed -- generally you'll look at clicks, but this could also be other types of conversions.

The calculator will spit out the confidence level your data produces for the winning variation. Then, measure that number against the value you chose to determine statistical significance.

14) **Take action based on your results.**

If one variation is statistically better than the other, then you have a winner. Complete your test by disabling the losing variation in your A/B testing tool.

If neither variation is statistically better, then you've just learned that the variable you tested didn't impact results, and you'll have to mark the test as inconclusive. In this case, stick with the original variation -- or run another test. You can use the failed data to help you figure out a new iteration on your new test.
While A/B tests help you impact results on a case-by-case basis, you can also apply the lessons you learn from each test and apply it to future efforts. For example, if you've conducted A/B tests in your email marketing and have repeatedly found that using numbers in email subject lines generates better click through rates, then you might want to consider using that tactic in more of your emails.

15) Plan your next test.

The A/B test you just finished may have helped you discover a new way to make your marketing content more effective -- but don't stop there. There's always room for more optimization.

You can even try conducting an A/B test on another feature of the same web page or email you just did a test on. For example, if you just tested a headline on a landing page, why not do a new test on body copy? Or color scheme? Or images? Always keep an eye out for opportunities to increase conversion rates and leads.