



HOW-TO DRAFT: A/B TESTING FOR SENDOUTS

This doc was designed to travel. Please feel free to share this link around with other progressive campaigners. You can download this (look under File menu top left) in several formats.

Disclaimer / request for your help	2
Summary	2
Who's doing it?	2
3 reasons to test	3
Impact/ Why do this?	3
When this might not work for you	4
Principles	4
What this requires (people, resources, etc.)	5
Setup steps / stages	7
Further resources	14
Attribution	15



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org



Disclaimer / request for your help

This is a work in progress that is meant to evolve over time with input from campaigners. At the moment, the tips and ideas here reflect the voices of the [contributors/reviewers listed below](#). We are always looking to add more voices of campaigners that have knowledge and experience on this topic. If this is you, please contact us here: blueprintsfc@gmail.com.

Summary

A/B testing involves testing campaign communications online with your supporters. Most of the guide focuses on testing email communications, but you can also test web page formats and social media responses. The idea is to make a change and try two approaches (A/B) in order to measure, through response statistics, which approach gets the most uptake from audiences. It can be seen as a way of optimizing campaign communications but is also viewed as a form of active listening for many groups who want to campaign in ways that are better aligned with the interests of their supporter base.

Who's doing it?

A/B testing has a long history in direct mail fundraising long before digital. However, running such tests through the mail was slow and costly so it had a pretty small niche.

In the digital age, advocacy networks like members of the [OPEN network](#), [Sumofus](#) and [Avaaz](#) refined and perfected this approach using petition-based platforms but now basically any advocacy group these days with big digital contact lists has worked some principles of testing into their practices.



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org

3 reasons to test

- 1) You're looking for an **immediate result** from a specific email so you can take on board learnings today. This is what is more commonly thought of in A/B testing, for example trying 2 email subject lines with a small group and sending the most popular to the whole list.
- 2) You want to **test something longer term**. This could be messaging, long term test - concepts or framings (e.g. human rights vs economic angle). This way you could take in learnings from a series of emails to inform a whole campaign.
- 3) You're looking to **test whole journeys**, e.g. 5 emails leading up to becoming an activist. You could set asks and frame them in a different way to see which approach works best in 5 emails over 6 months.

Impact/ Why do this?

- More impact
 - Leads to more people signing petitions, donating, calling, sharing
- Supporter-focused
 - Learn what messages resonate with your audience
 - Looks at real behaviour, not what people say would work with them (e.g. in surveys)
- Better decision making
 - Helps set goals and identify metrics that clarify success
 - Use of real-world data and behaviour that doesn't rely on perceptions or guesses on what **might** work
 - Avoids potential lost opportunities (opportunity cost) of doing one thing over another



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org



- It's easy
 - Lots of tools that make this easier than ever
 - Learning is fun!

When this might not work for you

If your organization is not willing to implement changes to campaign design and communications based on input from your supporter base, then you should not bother testing in the first place.

Also, if there no real organizational link between the person responsible for running tests and those responsible for creating campaign communications, then the testing will also be pretty much a waste of time.

If you only have a small list only more radical tests are likely to work for you - where you change bigger things. Ideally to test small differences, you'd want to trial each change with at least 5k people, then send the winner wider. If you only have an email list of 7k only more radical testing is likely to be useful, where you change lots of things. It won't be statistically significant, but it could give you indicative learnings about how your list responds.

Finally, if you don't have the capacity to easily message your members or track their interactions with you digitally, then testing would be hard to deploy.

Principles

The theory behind testing

1. **Test early:** We often use data *after* running a campaign in order to understand impact, but it's even better to use data to make campaign decisions along the way.
2. **Test often:** Set a goal to consider the potential of testing something, no matter how small, in every email you send. But only do this when appropriate! Occasionally testing is a waste of time - for example if you're emailing a small list of people about a one off event that won't inform anything in future there's



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org



no or be worth it.

3. **Report back results to your team** and if the results are inconclusive, try it again next time. You'll be surprised how easy it is to get hooked!

Even inconclusive results can be valuable - if there's no difference in response rates between a highly designed email with pictures to something more stripped back, for example, this tells you it isn't worth putting in extra time.

4. **Data-informed, not data-driven:** While "24 hours to save the world's saddest dolphin" may be the winning subject line, it's probably not the kind of brand your organization wants to project. Use data to inform your decisions but remember that data only tells half the story. Don't use data to replace instinct and qualitative feedback from your supporters.

What this requires (people, resources, etc.)

Tracking tools:

A/B testing usually requires access to digital response statistics. Most CRMs - Constituent Relationship Management platforms that manage member databases and comms (+ if you want [a more detailed CRM definition click here](#)) - ex. of some popular advocacy offerings include [Action Network](#) + [Engaging Networks](#) - come with emailer tools with built-in statistics dashboards for tracking email sendout performance and also have built-in A/B testing functions.

All encompassing testing-support software for website landing pages (the pages you're directing people to via email or elsewhere), such as [Optimizely.com](#) comes at a cost considered pricey by some groups (nonprofit rates exist) but can make the whole testing process a lot simpler and easier to manage, especially if it runs on several platforms.

Beyond or without platforms, A/B testing can be done through Google Analytics for website optimizations or through affordable email sendout software options such as



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org



[Mailchimp](#) (all email sendout software should do this really. It's worth changing your provider if it doesn't).

Social media content and promotions testing can be done through the statistics dashboards of most paid promotions and Facebook's Insights portal for Pages. For more advanced tracking of social media engagement, paid analytics options such as [Social Bakers](#) are available for \$120/mth.

Staff and culture

As important, if not more important than tracking tools, It's essential to have a culture where testing and failing, is ok. Lots of organizations are ok to test a new idea, but when it fails, they use it as an excuse to not test again.

In some organizations, testing becomes the sole domain of the digital department but for testing to work strategically, staff at all levels need to be involved. Ideally, staff involved in creating the content and the strategy behind it should be involved in the tracking the test.

New habits and practices need to be built around test management as well. For example, test results should be brought up in weekly meetings with staff at different levels of content and strategy involved. This way, the knowledge (value) gained from constant testing will reach throughout the organization.

Time

A quick A/B test on a planned message can take as little as 30 seconds to set up - for example if you're testing 2 different email subject lines with a send in Mailchimp, and you're just not sure which subject line works better.

Obviously, more involved testing projects, such as an audience consulting/listening exercise would be more time consuming.



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org

Setup steps / stages

The first step is to figure out what you want to find out, what success looks like, and how you're going to measure it.

Developing a hypothesis

Before each test you run, you need to run through this process to develop a hypothesis you're planning to test (remember high school science? Same kinda deal).

1. Goal

Start with the ultimate goal you're trying to achieve. e.g. We want to raise more money.

2. Big question

Break the goal down into a single big question – usually a 'What/where/why?' question. e.g. What channel is driving most of our online donations right now? (for the sake of this example, I'm going to say email).

3. Medium-sized questions

Break the big question down into smaller questions. You're trying to figure out the behaviour of your donors and supporters, so these will typically be 'how?' questions. e.g. How are donors able to access our donate page from our email channel? (Sample answer: by clicking a link in the email).

4. Smaller questions

You're nearly there – these are the questions you're hoping to answer with your experiment, and will typically be in the form of 'is/does?'. e.g. Does sending an email with a button link in it lead to more donations than an email with a text link?



5. Hypothesis

This is your time to turn that question into a statement. You absolutely have to be able to answer it with 'true' or 'false'.

e.g. Sending an email with a button link in it leads to more donations than an email with a text link (I can answer true or false to this! So we're good to go). You've identified the variables you're going to test - the things you will change - in the example comparing a text link with a button link to donate.

6. Metrics

So you've got your hypothesis! Now figure out what metrics you're going to use to test it out – this is absolutely crucial, and I'd suggest this is the time you talk to your tech and data people to make sure you can actually measure what you want.

For the example above, I'd look at primarily measuring this:

Total amount donated

But I'd *also* be keeping on eye on this stuff:

Total number of donations

Average donation (Total amount donated/number of donations)

So the good news here is that you now actually know what you want to test and how you'll measure it. The bad news is you've still gotta design and assess your experiment.

Setting variables to track

When testing, only test one variable at a time so you're comparing apples to apples.

Here are some ideas to help you get started:



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org

Blueprints for Change

Testing	Metric	Platform	What's needed
Subject lines	Opens	Mailer / CRM (member database software) sendout tool	Email set up, duplication, and sending
Best email content	Actions (not clicks – clicks don't necessarily equal more actions)	Mailer / CRM (member database software) sendout tool	Email set up, duplication, and sending Action page set-up and duplication
Page content	Engagement Conversions	Web page	Action page set-up and modification
Page views	Clicks via email	Web page	Action page set-up
Social media content	Engagement	Social media accounts	A Facebook page analytics account or instagram, snapchat twitter stats dashboard
"Growthiness"	New members	CRM (member database software), email signup lists	Checking an action's stats page (for New Activists)



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org

Sentiment	Feedback	Social media accounts, survey forms, email	Monitoring the info@ / general inbox Monitoring social media channels Survey/form set up
-----------	----------	--	--

Making sure you can access your data

If you don't have a one-stop dashboard for tracking test data, as is usually included in a CRM, consider create a spreadsheet that compiles different data sources and results. This will be essential for providing an 'at a glance' picture of results for team meetings and for tracking your overall approach.

You should also consider, and note in your spreadsheet, when you will be collecting that data. If putting results in from 1 email 2 hours after sending and then another email 10 days after it went out because you went on holiday, the two won't be comparable. To have comparable results you need to be strict on gathering them at the same time. If it's a long term test you may want to always look 3 days after sending an email - if it's an immediate test, like choosing the best subject line to send to a big email list, you'll want to look after a few hours.

Proper sample size

According to statistical norms around [statistical significance](#), the general rule is: The bigger the sample, the more reliable the results.

You're going to need to make sure you have a sample size big enough that you can draw conclusions with. The rule of thumb used in the industry is that 5k for each variable in an email is a good number.

However, a lot of smaller groups that want to get into testing may have email lists that total 5K or less. If test sendouts go out to segments that are portions lower than 5k, testing can still be useful to organizations provided that they run a series of tests over time and consider the trends they are seeing when viewing the results as a whole. In



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org



this case, it may take more time to extract useful messaging data but it will prove to be a valuable listening exercise nonetheless.

See also this [A/B Significance Test calculator](#): It tells you whether your A/B test is statistically significant and is useful for being confident that the changes you make will improve conversions.

Metrics that matter vs. “vanity metrics”

When running A/B tests, it may be tempting and most simple to track easily accessible statistics (email opens, website visits, “likes” on social media posts etc.) but these often do not generate the kinds of conclusions that can inform campaign strategy in any significant way.

To involve all levels of an organization in a conversation around testing, it’s important to be testing on statistics that clearly point out risks or benefits on issues of strategic significance to the org.

This is where the issues of “vanity metrics” vs. strategically significant metrics comes into play. While superficial metrics like opens, clicks and even list growth are easy to gather, they speak less to others than metrics more closely tied to desired outcomes, such as deeper member engagement and a bigger and more committed donor base, for example.

To go deeper here, [Mobilisation Lab and Citizen Engagement Lab put out a great report titled: Beyond Vanity Metrics in 2015 \(download here\)](#).

One interesting example coming out of the report above is Sumofus.org and their choice to measure MeRA (Members Returning for Action) as the number of existing Members Returning for Action on a monthly and quarterly basis.

Repeat testing

Best practice is running an experiment three times before making it your standard practice and testing something else. There could be a million reasons it worked once



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to’s and more info at: www.blueprintsfc.org



and not again – the Obama campaign famously saw a huge rise in donations through highlighting sections of text... that quickly wore off when they tested it again.

Don't get complacent - you are never done testing! Things can always change. It is worth coming back to testing things months down the line. Do learn and respond but times to send emails or preferred phrases can be fickle. Using capital letters and the word urgent in a subject line will increase opens, but if you do it too many times could have the reverse effect. Always revisit your assumptions.

Expanding the testing universe

If it looks good to a small universe, run the experiment again to a bigger audience before you go whole hog. It's a way of re-running the experiment, but getting a stronger result.

Leaving time for actions/responses

Give an email at least an hour - if you have a large list of 800-900k emails, you should have some useful results in this time. Even without having statistical significance, you'll likely see a trend that indicates a difference between the testing groups or not. The smaller your list, the longer it will take to gather statistically-significant data - if it's very small this may not be possible (see above for a stat significance checking tool) to collect data and make sure to track whether or not your predictions were accurate.

Analyzing your data

Action rate: This is the metric many org with large lists use to judge an email's performance. It's better than click rate – through the action rate, you can see how many people were driven to take action, but you can also figure out pretty quickly where in the chain something is going wrong if your email performs badly. If your click rate is high and your action rate is low, it usually means something's up with your landing page.

Amount donated (if it's a fundraising email):



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org



It's up to you whether you choose to look at average donation or total amount donated, but this can help you see if your email is inspiring people to give a higher or lower gift than normal.

Unsubscribe rate

Establish a baseline unsubscribe rate and just keep an eye on it. Unsubscribers aren't always bad anyway [link] but if they suddenly spike, you should definitely look into why.

A quick note on open rates

Open rates tell you how many people have opened an email, but a better indication of email performance is the percentage of the people who have opened it who have clicked a link and gone on to take action. Generally, look at open rates if your email is deeply underperforming – it could be an indicator of deliverability issues.

Otherwise, just keep your focus on the action rate - unless the open rate is key to what you're testing (e.g. if you're sending a newsletter to update people and don't want them to do anything else).

Keep track of your experiments

Keep track of your experiments, and share the findings with everyone on your team. Everyone works differently here, but keep some sort of testing spreadsheet, writing up the results, and ritually sending around new results and talking about them in meetings is a good way to share your testing wisdom.

Re-test your best practices

Every so often, go back and test something out again. You could be surprised.



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org

Further resources

Online tools:

- [A/B Significance Test calculator](#): Tells you whether your A/B test is statistically significant and is useful for being confident that the changes you make will improve conversions (but it can be expensive to reach statistical significant size and it isn't always worth it. Some data can be more useful than no data).
- Jon Lloyd provides an experiment checklist that you can download and physically check off as you're running your tests. ([downloads here as a pdf](#))
- Optimizely.com is a platform for tracking A/B testing experiments, [they have nonprofit pricing plans](#) but are still considered expensive.
- [Social Bakers](#) is a social media performance tracking tool that measure the rates of engagement with posts on all of the major platforms. Basic pricing to track 5 different accounts is \$120 USD/mth

Reports:

- [Beyond Vanity Metrics](#): Great 2015 exploration of groups seeking better metrics to test by [Mobilisation Lab and Citizen Engagement Lab](#)

Books:

- [The Moveon Effect \(2012\)](#) and the more recent [Analytic Activism \(2016\)](#) by David Karpf both explore the “culture of testing” as developed by digital-first campaigning groups and also how these practices have affected the larger advocacy sector in the U.S.



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org



Attribution

Input and resources for this how-to were provided by:

[Amara Possian](#) (ex Leadnow.ca) and [Jon Lloyd](#) (ex Sumofus.org),

This how-to was prepared by:

[Amara Possian](#), [Tom Liacas](#), [Natasha Adams](#), [Jon Lloyd](#)

Reviewed by: [Michael Silberman](#), [David Karpf](#)



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#).

For commenting permissions and other feedback, contact us at: blueprintsfc@gmail.com

Full library of how-to's and more info at: www.blueprintsfc.org