Research and Testing
Core Concepts

[Lab team member name here]
Why Run Tests? (and what are they?)

Stages in the Testing/Research Process

1. Concept Development
2. Design the Test
3. Collect Data
4. Analyze Data
5. Implement Insights into Programs!
Why Run Tests?
When you hear “test” or “experiment” – think “comparison”

What sets a formal test/experiment apart from just collecting data or “trying things out” is that you are consciously taking action and measuring the consequences via a careful comparison in order to learn what the effects were.
Testing is not...

- Public opinion polling/surveys
- Focus groups
- Media tracking
- Running programs and seeing what happens
- Just “experimenting” with wacky, out-there ideas
We test because...

• We want to know if \( x \) (our program) causes \( y \) (action on climate and energy issues!)

• We are not always great at guessing 😊

• Our work is **too important** and **too urgent** to spend time and resources in pursuit of strategies that are not moving us toward winning more policy victories and building more political power!
A/B testing: Obama 2012
Constantly looking for improvements, large or small, in every aspect of our digital operation

The testing mantra: test every element
Running tests in the background via personalized content
- Example: 300+ buckets for the “ask amount”

From: Barack Obama <info@barackobama.com>
Subject: My last campaign

David –
In a few days, I’ll be hitting the trail for my last campaign.

Everything we’ve accomplished in the past three years — and our chance to do so much more — is on the line.

What we do today will be a measure of whether or not we’re ready to fight for it.

Donate $5 or whatever you can before tonight’s fundraising deadline.

By pitching in before midnight, you’ll automatically be in the running to join me and George Clooney at a place on May 10th.

It’s not often I can get away from work, so I look forward to spending a fun evening in L.A. with a couple supporters like you.

“We basically found our guts were worthless.”
Q: Why does identifying your best-performing content and evaluating your programs matter?

A: Because what you learn can have big consequences.

One progressive testing firm found top performing ads often 2x to 7x more effective than average ads.

Another’s testing has found that about 1 in 5 ads backfire (persuading audiences in the opposite direction).

A review of tests in general election campaigns found zero effect from persuasion tactics.*

*Kalla & Broockman, 2018
How to motivate climate action...in Texas!

You’re not alone. A growing number of Texans (now 71%) agree we should do more to fight climate change, including a majority of Texas Republicans. However, U.S. oil and gas production is poised for an expansion that will be unparalleled.

2 percent reduction in cost per engagement from added social norm language.
Stages in the Testing/Research Process
Steps in the Testing Process

**Concept Development**
Clear idea of what you want to test, prediction of what will work and why it should work.

**Design**
Create variations in program to be tested, carefully varying only the elements (e.g., subject line) to be tested.

**Collect Data**
Send email, buy Facebook ads, etc., making sure versions are randomly distributed among participants.

**Analysis**

**Implementation**
Integrate results into new activity or revise existing program.
FULL-STRENGTH
testing got you feeling nervous?

That’s cool!

A great quick and easy way to get started with testing is to experiment with subject lines in your email sends.

You can do this without too much planning, and most email platforms have a built-in testing feature that makes it simple.
1. Concept Development
What is a question that, if you knew the answer today, would lead you to make decisions differently tomorrow?
Get from *general* to *specific* research questions

- **Not so good**
  - How do we fight Trump?”

- **Better**
  - “Does referring to Trump fire people up?” (specific)

- **Even Better**
  - “Does a Facebook ad referring to Trump generate more public comments from our base than one referring to utilities?” (more specific)

- **Best**
  - “A Facebook ad referring to Trump will generate more public comments from our base than one referring to utilities, because anger and fear are motivating.” (hypothesis with theory)
General testing principles

Play with what you’ve got

• Instead of investigating a totally new idea, test a variation to an existing program.
• Ensures results will be actionable + allows you to compare results against an existing baseline (“we improved 25%”).

Go big

• Experiment with substantial changes, not small tweaks.
• Tiny differences (in wording, etc.) rarely make a (significant) difference.

Focus on the relevant + actionable

• How will your work change based on the results of your test?
• Quickly showing concrete impact builds buy-in for testing across the organization

Test on meaningful metrics, not vanity metrics

• Beyond clicks and size, focus on impact.
• How do you know your theory of change is working?
2. Design
Two key aspects of rigorous testing design

Careful variation

Random assignment

Phonebank with us!

Lobby with us!
Greater Cleveland participants each were randomly shown one message variation and asked about their support for the 100% clean energy plan.

**HEALTH FRAME**
**VS.**
**JOBS FRAME**
**VS.**
**LEADERSHIP FRAME**

Potential supporters

Survey question

- I oppose this 100% clean energy plan
- I support this 100% clean energy plan
A/B Testing

50% visitors see variation A

21% Conversion

50% visitors see variation B

38% Conversion
California EJ Alliance A/B Message Test

Research Question: *Which framing strategy would drive more email actions?*

Framing the campaign goal as...

**Fight Back Against Big Oil**

Framing the campaign goal as...

**Protect Community Health**
[A] Fight Back Against Big Oil

Environmental justice communities suffer from a range of pollution sources all at once. On top of extreme heat, poor air quality, or access to housing or jobs, all of these climate impacts are and will increasingly be felt in the same communities that have the fewest resources to adapt. California’s families, workers, and children are held hostage by the power of Big Oil and must be protected from dirty smog clogging their lungs everyday.

[B] Protect Community Health

Environmental justice communities suffer from a range of pollution sources all at once. On top of extreme heat, poor air quality, or access to housing or jobs, all of these climate impacts are and will increasingly be felt in the same communities that have the fewest resources to adapt. The health of California’s families, workers, and children must be protected from dirty smog clogging their lungs everyday.
A/B tests vs. Randomized Controlled Trials

A/B Tests

• Use-case: when you want to find which version works better
• Great for quick gut-checks or detail work

Randomized Controlled Trials

• Use-case: when you want to find out why something works better
• Better for digging up mechanisms

Phonebank with us!
Lobby with us!
Control group
Message(s)
RCT example

“Control”

“The Trump administration’s pipeline plan will lead to a calamity of oil spill disasters across the country. Add your public comment now!

John Doe and 20 Others

11 Comments

“Picture this” text

“Picture this landscape without the pipeline. Find out what you can do to make it a reality.

John Doe and 14 Others

5 Comments

Trump image

The Trump administration’s pipeline plan will lead to a calamity of oil spill disasters across the country. Add your public comment now!

John Doe and 33 Others

19 Comments
IF this had been an RCT...

**[Control]**

**Fight Back Against Big Oil**

Environmental justice communities suffer from a range of pollution sources all at once. On top of extreme heat, poor air quality, or access to housing or jobs, all of these climate impacts are and will increasingly be felt in the same communities that have the fewest resources to adapt. The health of California’s families, workers, and children must be protected from dirty smog clogging their lungs everyday.

**[Treatment]**

**Protect Community Health**

Environmental justice communities suffer from a range of pollution sources all at once. On top of extreme heat, poor air quality, or access to housing or jobs, all of these climate impacts are and will increasingly be felt in the same communities that have the fewest resources to adapt. The health of California’s families, workers, and children must be protected from dirty smog clogging their lungs everyday. California’s families, workers, and children are held hostage by the power of Big Oil and must be protected from dirty smog clogging their lungs everyday.
3. Collect Data
“Data is how you listen to people at scale”

—FB analytics lead
What outcomes to test?

- **Vanity metrics**: Easy to measure, but not meaningfully linked to your strategy or theory of change
  - Examples: *Clicks, views, list size*

- **Meaningful metrics**: Those that inform decision-making towards accomplishing your goals
  - Examples: *Identification with your organization, leadership upladdering rate, legislative co-sponsors*
Measuring People Power

A global survey of hundreds of change-makers

- Many organizations still emphasize vanity metrics (such as list size)
- But innovators are measuring depth of engagement
- Some are developing custom measurement of their people power to fit their theories of change
What do you want to know from your test?

• **What people think.** Use survey-experiments & respondent pools

• **What people do in the real world.** A/B test with email, Facebook, your site, your programs.

• **Both!** Conduct hybrid tests with a survey and a real-world component
Measuring what people think

- **Examples:** Persuading, informing, raising importance, counter arguing, reframing
- Harder to measure than behavior
- **Methods**
  - Placebo/no message controls
  - Pre-post measurement
  - ‘Lab’ pre-testing (survey pools, e.g., Swayable, Google Surveys)
  - “Brand lift” surveys (within-platform)
  - Sentiment analysis of posts and comments
Tools for testing persuasion
Brand lift studies & survey-based platforms

YouTube/Google Brand Lift

Google consumer surveys

Holdback Group

amazon mechanical turk

SWAYABLE
Measuring what people do

- **Examples**: actions, open rate, click rate, shares, likes, comments, traffic, sign ups, conversions, event attendance, policymaker calls
- **Easy to measure within platforms, CRMs, web**
- **Methods**
  - A/B tests
  - Dashboard stats
  - Tracking/conversion pixels
  - Referral link codes
  - Paper & pencil tracking
Field tests measuring actions

- Facebook
  - Popular testing platform, but design logistics can be tricky
- Email
  - Most bulk email platforms have built-in testing tools
- Online advertising (Google Adwords, etc.) acquisition cost
- Testing of websites/landing pages via tools like Optimizely, Google Analytics, etc.
- Any systematic large scale campaign activity
  - Embedding tests in phonebanks/robocalls (e.g. patch-throughs), canvass scripts, volunteer engagement programs, etc.
4. Analysis
“Analysis” at its most fundamental just means systematically examining your data and drawing conclusions from it. And it doesn’t always have to be super complicated!

Thoughtful analysis is important because simply plucking out an arbitrary number (We got 100 Likes in the first 10 minutes!) or anecdote can lead to misleading conclusions and poor decision making.

You’ll use different analysis techniques for quantitative data (numbers, like action rates, polling, experiment results) than qualitative data (text, like focus groups or stakeholder interviews).
Data analysis 201

• Statistical analysis is used for quantitative data
  • **Descriptive statistics**: Simply summarizing the data, such as averages, counts, minimums, maximums
  • **Inferential statistics**: Are results reliable? Such as margins of error, statistical significance tests, regression

• Avoid over-generalizing results: The uniqueness of every research project limits how broadly you can apply the findings
  • *Would the results be different with another constituency, at another time, on another campaign?*

*Inferential statistics* tell you if your data contain a meaningful signal about the world, or are more likely just due to random chance.

*Example: statistical tests tell you if a higher click rate means a subject line was truly better, or just a lucky blip*
Analysis options

**DIY**
- Paste & click online calculators (e.g., M+R’s Toolshed), Excel’s Analysis Toolpak*, statistical software (*R* is free), online word cloud generators

**Platform-based**
- Tools such as Optimizely, and many CRMs automatically analyze

**Outside help: The more sophisticated consultants can do this, including…**
- Partners and vendors such as the Analyst Institute
- Collaborate with the Yale Program on Climate Change Communication’s Facebook testers
- Your friendly neighborhood social scientist!

*https://bit.ly/2HUxLz9
What is a question that, if you knew the answer *today*, would lead you to make decisions differently *tomorrow*?
Implementation tips

- **Buy in:** Involve everyone early who will need to sign-off and implement.

- **Compare to a baseline:** Design experiment so it can show an X% improvement over how you already do things.

- **Test grounded questions:** Before running test, ask yourself, “how would we do things differently if we knew the results?”

- **Be ready for failure, and improve:** Only 26% of average Optimizely tests worked (but 77% of those following best practices worked).

- **Promote!** Share test write-ups within your organization and with the Climate Advocacy Lab.
BUILD IT

MEASURE IT

LEARN FROM IT