

THE AGENCY STANDARD

THE TESTING AND SCALING PLAYBOOK

One of a series on building outbound that books meetings.

REVENUE.INC × CLOSE3X

PRIVATE CIRCULATION

THE TESTING AND SCALING PLAYBOOK

“There are no universal benchmarks. Only yours.”

A REVENUE.INC × CLOSE3X FIELD GUIDE · TESTING

ESTABLISHED ON 10,000,000 SENT EMAILS MONTHLY · TWO HUNDRED CLIENT DESKS

What this is. *This is the field guide to knowing whether a campaign is working and what to do about it. It opens with the uncomfortable part: there are no universal benchmarks, because every business has different economics, and a reply rate that is excellent for one is a disaster for another. From there it covers the three things worth testing, in the order they matter, the rules that keep a test honest, how to read heat signals rather than vanity numbers, and when a winner has earned more volume.*

01 NO UNIVERSAL BENCHMARKS

Before you test anything, understand that benchmarks are not universal. Every business has different economics, so what counts as strong performance for one is a disaster for another. There is no single reply rate, meeting rate, or close rate that is good everywhere, and the fastest way to make bad decisions is to judge a campaign against numbers that belong to a completely different kind of business.

The single biggest factor is LTV. A company charging \$20 a month is playing a completely different game from one sitting on \$250,000 in LTV, and the two need to be tested in completely different ways. The economics decide how much performance you need before the campaign makes money.

For low-LTV businesses, you need strong performance to make money, because there is almost no margin to absorb a weak campaign. Every reply, every meeting, and every close has to be efficient, so your bar for what counts as a working test is high.

For high-LTV businesses, you can afford far worse performance and still make a fortune, because a single closed client is worth so much that even an expensive, inefficient campaign pays for itself many times over. The same numbers that would sink a low-LTV business are perfectly fine here.

So calibrate your expectations to the LTV before you judge any number. Everything that follows is a set of rules of thumb, not hard laws. Use them as a starting point, then adjust them to the business in front of you.

BENCHMARKS ARE NOT UNIVERSAL

The lower the LTV, the better your performance has to be. The higher the LTV, the more inefficiency the business can absorb. Judge every number against the economics behind it.

02 THE THREE WAYS TO TEST

There are only three things you can ever test on a campaign. Deliverability, targeting, and messaging. Everything you could possibly change falls into one of these three buckets, so every valid test is just a change to one of them.

It helps to think of them in order of size. Deliverability is the smallest lever, targeting is much bigger, and messaging is where most of your wins will come from. The sections that follow break down each one, what you can test inside it, and how much of your attention it deserves.

THREE LEVERS

Deliverability, targeting, and messaging. Every test you run is a change to one of these three, nothing else.

03 DELIVERABILITY

Deliverability is the smallest of the three levers, so don't over-invest here. It matters, but it's rarely where the big wins come from, and it's easy to burn weeks chasing tiny differences that don't move the business. Give it a light test, learn what works, then spend your real energy on targeting and messaging.

What you're testing is the type of email account you send from, the ESP. Sometimes you'll get better replies from Google, sometimes from Microsoft, sometimes from plain SMTP accounts. It varies by market and by list, so it's worth trying all three, but keep it light.

1. *Google*. Gmail and Google Workspace sending accounts.

1. *Microsoft*. Outlook and Office 365 sending accounts.

1. *SMTP*. Standalone SMTP sending accounts.

DON'T OVERDO IT Test Google, Microsoft, and SMTP, then leave it alone. Deliverability is real, but it's the smallest lever of the three.

04 TARGETING

Targeting is a much bigger lever than deliverability. There are four big variables you can test, and then one that's even bigger than all of them, which is how you get the leads in the first place.

Start with the four core variables. Change one at a time so you can see what the change did.

1. *Industry*. The vertical the company operates in.

1. *Company size*. The employee range, like 1 to 50 versus 50 to 200.

1. *Job title*. The seniority and role of the person you're reaching.

1. *Location*. The country or region the company is in.

How you get the leads

Beyond the four variables, the single biggest thing you can test is where the leads come from in the first place. A perfectly written email sent to the wrong list will always lose to an average email sent to the right one. The way to think about it is to divide every lead source into three priorities, from warmest to coldest, and test across them. Here are a few from each, not the full list.

Priority 1, your own audience. The hottest leads you'll ever email, people who have already interacted with you in some way. They already know who you are, so they convert at the highest rate. Always start here.

- Past clients and closed-lost prospects.
- Newsletter, YouTube, or podcast subscribers.
- People engaging with your LinkedIn posts and profile.

Priority 2, engagement signals. Warm prospects who haven't reached out yet but are showing buying signals, either through the tools they run or a recent change in their business. They just need a nudge.

- Companies using a competitor's tool, or a complementary tool that implies they need yours.
- Recently funded companies, where budget is moving right now.
- Companies hiring for roles relevant to what you sell.

Priority 3, scalable cold outreach. The coldest and lowest intent, but the way you fill the pipeline fast when the warmer sources run dry. There's no signal here, just a tight fit on paper, so get the profile right or the whole list is wasted.

- Firmographic filtering by revenue, employee count, industry, and location.
 - Lookalike audiences modelled on your best existing customers.
 - Job title and seniority filtering inside target accounts.
-

START WARM

Work your own audience first, then engagement signals, then cold. The warmer the source, the higher the conversion, so only scale into cold once the warm ones run dry.

05 MESSAGING

Messaging is where most of your testing time should go, because it's the biggest lever of the three. And within messaging, you always start with the offer, because the offer is everything. A great offer with average copy beats a weak offer with perfect copy every single time, so test that first, then work your way down the list.

These are the things you can test, roughly in order of impact.

1. *The offer.* By far the most important. Test different offers before you touch anything else.
 1. *Pain points.* Which problem you lead with and how sharply you name it.
 1. *Value proposition.* How you frame the core benefit they walk away with.
 1. *Social proof.* Which proof points and results you put in front of them.
 1. *Call to action.* What you ask them to do at the end.
 1. *Subject lines.* What gets the email opened in the first place.
 1. *Email length.* Short and punchy versus longer and more detailed.
 1. *Personalization.* How much you personalize and what kind you use.
-

OFFER FIRST

The offer is everything. Test it before anything else, because no subject line, CTA, or clever line will save a weak offer.

06 RULES OF THUMB FOR TESTING

A few simple rules keep your testing clean, valid, and moving. The most important idea underneath all of them is what makes a test valid in the first place. A valid test means you change exactly one of the three levers, the messaging, the targeting, or the inboxes, and hold everything else steady. Change two things at once and you'll never know which one moved the number.

With that in mind, run your testing by these rules.

1. *Always have at least 3 campaigns running.* At any given time, keep a minimum of three live, each with a different valid test.
1. *At least 3 variants per step.* Every campaign should have at least three variants for every step in the sequence.
1. *Cut the dead variants.* Turn off any variant with under 1 lead after 1,000 sends. It has had its chance.

WHAT MAKES

A TEST VALID Change only one lever at a time, messaging, targeting, or inboxes. Change two and you'll never know which one moved the number.

07 HOW TO OPTIMIZE, HEAT SIGNALS

Once campaigns are running, you optimize by following the heat. Heat signals are the patterns that show you where the money is coming from, and usually that's the same place you're getting the most meetings. So you look at where your replies and meetings concentrate, then double down hard on it.

These are the things to watch.

- Company size.
- The job title of the people responding the most.
- The industry.
- The location.
- The messaging that keeps landing.

Even if each of the five variables is only 20% better, and not even 2x, stacking all five is 1.2 to the power of 5, which comes out to about 2.5x overall. Small edges on every layer compound into a campaign that performs multiples better than where you started, so you chase every edge, not just the obvious ones.

One important caveat. Heat is a proxy for revenue, not the goal itself. The rule of thumb is that where you get the most meetings is where you get the most revenue, so you hammer down there. But if the reality is different, if all of your actual sales come from, say, marketing directors at agencies in Europe in the 200 employee range on one specific offer, even though they look worse on reply rate, you follow the sales, not the meetings. Always double down on where the money lands.

FOLLOW THE MONEY

Double down on the segment that produces revenue. Meetings are usually the signal, but if the sales come from somewhere else, follow the sales every time.

08 WHEN TO SCALE

You scale a campaign once it's proven it can produce results cheaply, not before. Scaling something unproven just means you lose money faster. The benchmark is simple. When it takes fewer than 500 emails to get one positive reply, the campaign has earned the right to scale.

When you do scale, don't pour everything into it at once. Put the proven winner at around half of your total sending volume and treat it as your baseline, the control that everything else gets measured against. Use the other half to keep testing new variants, so you're scaling what already works while you keep hunting for the next edge. That way you grow the winner and keep improving at the same time, instead of betting the whole operation on a single campaign.

THE SCALE THRESHOLD

Fewer than 500 sends per positive reply means it's ready. Run the winner at about half your volume as the baseline, then keep testing everything else against it.

THE LAST WORD

Most teams test the wrong layer and scale on the wrong signal. Fix deliverability before targeting, targeting before messaging, and change one thing at a time or you learn nothing. Then judge on meetings and revenue, not opens. When something wins, scale it slowly enough that you can tell whether it is still winning. Patience here is cheaper than a burned domain and a quarter of bad data. If you would rather have this run for you than run it yourself, that is the conversation this was meant to start. Book a working session at close3x.com/demo or revenue.inc/demo, and we will map your setup and show you exactly where the leaks are.

REVENUE.INC**THE AGENCY**

Fully managed outbound for B2B. We fill your calendar, you close the deals.

revenue.inc/demo

CLOSE3X**THE TOOL**

Know who answers before you dial. Every number scored before your reps call.

close3x.com/demo

WRITTEN AND RUN BY CHRISTIAN OLAND

**linkedin.com/in/c-oland
youtube.com/@Christian_Oland**

FOLLOW THE COMPANIES

**linkedin.com/company/rev-enue-inc
linkedin.com/company/close-3x**