MarketDial

In-store A/B testing:
Operational strategy handbook

——

Contents

▌Introduction

We've interviewed 20+ organizations with varying levels of testing maturity, fleet sizes, and governance models to find out what testing processes produce optimal outcomes. Based on those interviews, we have compiled a guide of the best practices and tactics retailers use to organize and coordinate the testing process from start to finish.

Those who are new to the testing process do not need to implement all these tactics right away since they may be running just a couple tests. However, as organizations increase testing breadth and volume, we find that having tactics in place is key to ensuring success from testing. We'll note the items that are good to adopt early on and others that come at a later stage.

Testing is a joint, company-wide effort. The most successful companies have buy-in throughout the organization, enabling cohesion in design, implementation, and decision making across teams.

As your organization continues to grow in testing maturity, it's helpful to define ownership and implement regular touchpoints and expectations throughout the testing process to instill good testing practices and enable coordination across teams. Test implementation is most effective when included as part of your strategic and operational planning, with processes, timelines, and roadmap in place.

The following sections will address operational strategies and best practices for each step of the testing process to ensure success.

01 Designating key roles

Identify people and teams that will be impacted and that need to be involved.

02 Test idea generation

Define the business problem and proposed options for solving that problem.

03 Test design

Choose the right stores for your test that are reasonable for implementation and ensure a good read.

04 Test implementation

Store teams implement the initiative in select test stores.

05 Test results & review

Best Practices

Designating key roles

The first step of kicking off any project – testing or otherwise – is to identify the people and teams that will be impacted and that need to be involved. There's no one-size-fits-all approach to assigning testing roles.

In general, we have identified five key personas that play a primary role in the testing process. These personas are defined based on their responsibilities in the testing process; however, across different organizations, the same person can take on multiple roles. It's less important how many people are involved and more important that key responsibilities are assigned to an owner who knows best practices for their designation.

◇ **Ideator**: Comes up with test idea

◇ **Project manager**: Plans and manages test implementation and tracking

◇ **Testing Subject Matter Expert (SME):** Designs and analyzes the test, builds business case for project rollout

◇ **Store team:** Implements the operational aspects of the test in stores

◇ **Decision maker:** Determines whether a test initiative should be rollout more broadly based on test results and other relevant factors.

Best practices

Test idea generation

In some ways, the test idea generation step is the most exciting step of the testing process; teams have the opportunity to be creative, innovate new ideas for in-store changes, and enhance the organization's strategic goals.

Without implementing best practices, however, test ideas can break down. Teams may rush to implement their ideas without thinking through how this may impact the test itself or changes being implemented by other teams. Some may try to test too many test ideas at once, without enough stores to run them all accurately. Others may test changes without fully defining goals or aligning them with company priorities. If teams don't have a clear vision of the value of testing, they may also roll out changes without testing at all. Any of these errors can lead to major losses or setbacks.

To avoid these mistakes, make sure the right people are in constant communication, testing the right initiatives in the right order. Follow these best practice guidelines for establishing a winning test idea generation and prioritization strategy.

01 Set regular cross-functional touchpoints

Given the cross-functional nature of testing, it's important to set regular touchpoints with relevant stakeholders. If done consistently, the meetings help keep testing top of mind for teams across the organization. In addition, touchpoints allow the organization to collect all ideas and prioritize them, setting a collective schedule for testing. Consistent coordination, training, and communication also ensures each team is adhering to best practices and allows teams to share learnings and grow together.

For organizations that are newer to testing, we've found a good first step is to set up a **testing interest group**, a cross-functional group of ideators and Testing SMEs that meets on a monthly-to-quarterly basis to discuss testing topics. This group helps promote visibility into testing across the organization with topics ranging from recent business decisions driven by testing to upcoming business priorities and test ideas.

As testing volume increases, it often becomes necessary to set up a **testing steering committee** that also meets on a monthly-to-quarterly basis to help teams coordinate testing schedules to avoid interference. At a glance, a steering committee may look very similar to an interest group, but the goals are a bit different. A steering committee is in charge of prioritizing test initiatives and setting testing priorities and goals. The group is most effective when it includes both a) cross-functional leadership with decision-making power and b) testing SMEs who can act on the decisions coming out of the committee.

02 Ensure tests align with the organization's strategic goals

Testing is most valuable when it drives the direction and decisions of the business. A very simple way to ensure your testing program aligns with your company's priorities is to map each test to the strategic area it falls under. Most organizations have regular sessions to align on annual or quarterly goals of the company, testing or otherwise. These goals should be widely distributed across the organization; each test that is run should be mapped back to the goal it helps meet – either in reporting or during cross-functional touchpoints. This practice generally helps align team efforts with company efforts.

Thoughtful strategic planning will create cohesion across the organization, enabling you to:

Filter and set testing priorities

Not every initiative needs to and/or should be tested; this process helps sift out the less important initiatives and ensures that attention is given to strategic initiatives that align with company goals.

Keep stores happy

Your size and store team capacity are limited. Careful filtering and prioritization helps ensure that you aren't overwhelming your stores with too many initiatives to implement at once.
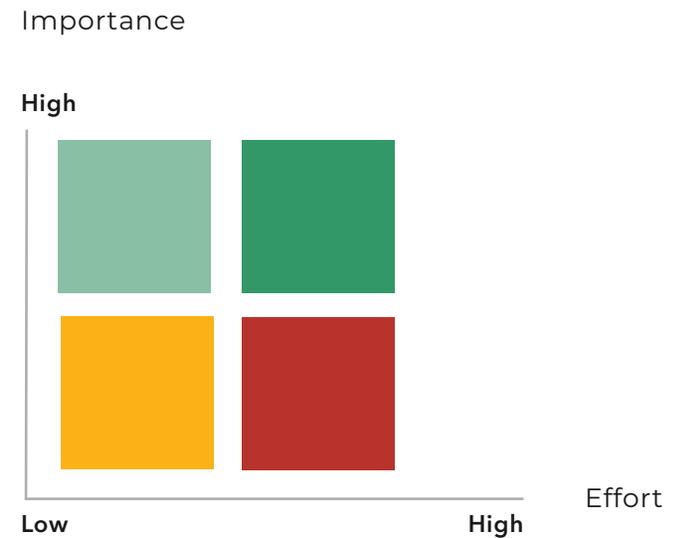
Promote a clean test read

Effective organizational test planning prevents overlapping tests from impacting each other's results.

03   Make prioritization a priority

As your organization matures in testing, it may reach a level of testing volume where there are not enough locations in the fleet to test all your ideas at once. While this is a great problem to have, it does become essential that you set a roadmap for testing. Depending on how your organization is structured, testing prioritization can be done at a team-level or at a company-level.

Similar to how engineering teams score and prioritize tasks to set an engineering roadmap many retailers approach test prioritization by scoring each test idea based on two factors: importance and effort. **Importance prioritization factoring** is a measure of how much value the initiative can bring to the business, and is often determined based on estimated impact to a key business focus (e.g., store financial performance, customer experience, operational efficiency). **Effort prioritization factoring** is a measure of complexity and resources required to implement the initiative—both for testing and rollout—and is determined based on cost factors (capital expenditures, labor). High impact tests are given highest priority on the testing roadmap, while low impact/high effort tests are deprioritized.

Importance

High

Low   **High**   Effort

High importance/high effort tests are prioritized while low importance/high effort test are deprioritized. Low importance/low effort tests are individually assessed and may move to higher priority as needed.

Best practices
Test design

Once you've aligned on and prioritized the tests your organization wants to run, the next step is to make sure you design a statistically valid test while simultaneously balancing business constraints. Instilling best practices will promote strong test design across your organization, enhance training and advoacy to strengthen test outcomes, and help secure buy-in from stakeholders.

When beginning to test, the test design process can feel daunting. It helps to break each part down into manageable phases. There are three basic test design decisions to make from the outset: 1) Choosing the optimal number of test stores and test weeks, 2) selecting the ideal set of sample stores to test in, and 3) matching the test stores to the most similar control set. The MarketDial platform helps automate the science behind these decisions, making test design both easy and accurate.

In addition to using the right software for test setup, it is essential to create alignment on test design throughout your organization. While that may feel daunting, following a few simple guidelines can help get everyone on the same page.

Consider starting small

Many retailers find it useful to implement a phased testing approach, starting small at a couple of stores to first establish operational feasibility and slowly expanding testing to measure ROI and rollout efficacy.

Oftentimes stakeholders struggle to understand the importance of testing with a large sample, thinking it's too risky or too costly. Implementing tests through this phased approach can help stakeholders work up to larger samples via iterative testing.

Small tests help develop test implementation strategies, gathering

qualitative feedback from store associates. Medium-sized tests in a single market also tend to be easier to implement and can provide directional guidance on implementation challenges before proceeding further.

If small and medium tests have gone well, expanding to a larger test is usually straightforward, utilizing a fully representative treatment group that can accurately estimate the rollout ROI in advance of a final rollout decision, providing useful decision intelligence for high-risk, high-cost changes.

	Small 1-2 sites	**Medium** 2-10 sites	**Large** 10+ sites
Decision	Can we execute the test?	Do we have the right hypothesis?	What's the expected rollout ROI?
Goal	Prove operational feasibility	Directional read on the test	Clear read on the test with high confidence for rollout decision
Example	Install EV chargers, single store	Install EV chargers, tested in an entire market	Install EV chargers, tested in a representative sample of stores across markets

Consistently educate stakeholders on best practices

A common setback with test design occurs when stakeholders or ideators rush to test ideas, then measure retroactively. Stakeholders will often self-select treatment stores themselves without considering how this can introduce noise or statistical inaccuracies. These treatment stores may consist of too small of sample or reflect a very biased representation of the store fleet (e.g., stores close to HQ, top performing stores). To prevent this from occurring, institute the following best practices at both macro and micro levels:

Set organization-wide requirements around test design
Ensuring test design protocols are standardized across an organization enhances trust in the outcomes and gets everyone speaking the same testing language. Leadership can start by requiring their teams to work with the testing SME up front when selecting treatment stores for their tests.

Provide constant education testing best practices
Hold testing education sessions during regular cross-functional testing touchpoints (testing interest group, testing steerco). Consider one-on-one education sessions with stakeholders regarding their tests. Utilize this time to clearly explain to stakeholders the implications of poor test design on test results. Encourage stakeholders to engage with the testing SME earlier to help with testing design going forward.

Test Results

Lift: **Confidence:**

+0.1%   2.6%

Key Takeaways

- Overall, results showed not significant lift

- We would need to see a much higher lift to have confidence in results, for example:

Lift:	Confidence:
+0.1%	2.6% (Poor)
+5.0%	65% (Fair)
+6.0%	75% (Good)
+9.0%	90% (Best)

- Alternatively, expanding the test to an additional 50-100 doors would lower the lift threshold to reach high confidence in test results

Presenting information to stakeholders in formats such as this can educate them on the necessity of high-caliber test design and help them understand next steps in the design process.

Best practices

Test implementation

Correct test implementation is a huge part of the testing process but often the part of testing that retailers feel they have the least visibility and control over.

Getting store teams onboard and managing implementation resources are two big challenges that come with testing at scale. Luckily, by following a few simple steps, retailers can ensure this part of the testing process runs smoothly.

Step

01 • **Ensure your test stores are operationally feasible**

Step

02 • **Build a detailed implementation plan for each test**

Step

03 • **Be thoughtful about the timing of each test**

Step

04 • **Require business signoff of implementation plan**

Step

05 • **Put your plan to action and monitor progress!**

Test implementaton process

Implementation will feel less daunting if broken into bite-size pieces. A step-by-step framework can guide you seamlessly through the process one step at a time.

Step
01

Ensure your test stores are operationally feasible

Generate a list of hard constraints that are essential for each test such as stores having the applicable store banner, layout, or products. Then be sure to vet all the selected treatment stores to ensure they satisfy all the restraint criteria.

Step
02

Build a detailed implementation plan for each test

Create a plan that walks all invested parties through the materials and processes that will be required. Be sure to consider each of the following:

- ◇ Impacted stores
- ◇ Store comms
- ◇ Timelines
- ◇ Trainings
- ◇ Labor impact
- ◇ Equipment delivery/installation plans

Test implementaton process

Implementation will feel less daunting if broken into bite-size pieces. A step-by-step framework can guide you seamlessly through the process one step at a time.

Step
03

Be thoughtful about the timing of each test

Test implementation takes time and effort, so make sure you give your stores sufficient time to plan and execute the changes. Most retailers plan on 2-8 weeks, depending on the size of the change.

Whenever possible, align your test start dates with store ops schedules. For example, many retailers change up their in-store planograms on a quarterly basis. If you're testing planogram strategies, align those test dates around when the planogram resets typically occur.

Avoid implementing high-effort tests during seasonally busy periods such as holidays. This tactic has the dual benefit of:

1) making sure your store teams aren't overwhelmed, and

2) reducing noise when reading results of your test.

Test implementaton process

Implementation will feel less daunting if broken into bite-size pieces. A step-by-step framework can guide you seamlessly through the process one step at a time.

Step

04

Require business signoff of implementation plan

Once the planning is complete, obtain final signoff from the business before going live to confirm that it's operationally feasible. Having clear approval ownership and clear standards/templates around implementation plans can significantly streamline the process.

Some retailers create a centralized **capacity planning team (CPT)** that approves stores, timelines, and implementation plans for every change that takes place in stores, whether it be testing or rollout. Having a CPT ensures 1) test goals align with company priorities; 2) test stores and dates are operationally feasible; 3) test design aligns with test goals; and 4) the implementation plan is detailed and reasonable.

Under the guidance of a CPT, every team involved with testing can submit an implementation proposal for approval that includes:

◇ What changes will be made

◇ What are the goals for the changes

◇ Impacted stores and dates

◇ Implementation plan

Irrespective of whether or not you implement a formal approval team, the process of obtaining approval can create greater test continuity across all levels of the organization and among personnel involved.

Test implementaton process

Implementation will feel less daunting if broken into bite-size pieces. A step-by-step framework can guide you seamlessly through the process one step at a time.

Step

05

Put your plan to action and monitor progress!

Continue to work closely with store and district managers to ensure correct implementation. If possible, leverage data to track implementation progress, and consider occassional store visits to confirm correct test implementation.

If your test implementation process is rocky, keep in mind that poor test implementation can inform future testing. You can always tweak the implementation plan and try again! This is an added benefit to the phased testing approach we described above: you can learn from mistakes in earlier phases to develop the right implementation playbook for broader rollout. Additionally, when you actually roll out a test initiative, chances are that implementation won't be perfect at all stores. Test results that include the noise of poor implementation at some stores can be more reflective of the impact of broader rollout.

Best practices
Test results & review

Upon successful completion of tests, organizations are now equipped with the information needed to make effective deicisons. But your role as a testing advocate within the organization is just beginning.

You can utilize test results and rollout decisions to:

◇ Provide consistent decision intelligence across the organization;

◇ ensure testing results are getting leveraged when relevant in decision-making, and

◇ increase the visibility testing to promote more testing.

When it comes to weighing decision intelligence for making final decisions, many retailers find it valuable to create a decision committee. These groups oversee decisions related to changes with high capital expenditures. Decision committees can be cross-functional groups or be independent from all other parts of the organization.

Leadership required every capital expenditure over $1M be tested and validated through MarketDial before rollout. The strategy upheld high testing standards across the organization

and ensured consistent decision criteria were implemented company-wide.

Even in the absence of a formal decision committee, maintaining consistent test report templates and decision criteria across the business is emphasized to ensure uniform processes. A monthly or quarterly newsletter can be a great way to provide high-level updates on the testing pipeline and share recent test results. Keeping the organization informed of recent decisions based on test results can emphasize the importance of testing and enable everyone to celebrate its successes.

Developing a decision-making strategy

A large grocery chain wanted to improve decision-making and prioritization across the org. Their decision committee determined that every initiative over $1M needed to be tested through MarketDial, creating a company-wide, consistent strategic standard for data-backed decisions.

Moving forward

Building on test learnings

The successful implementation of tests in a retail organization hinges on a systematic approach that encompasses meticulous planning, clear communication, and an organizational commitment to data-driven decision making.

By defining clear objectives, prioritizing test ideas, and designing tests that are statistically valid and aligned with strategic goals, retailers can lay the groundwork for effective experimentation. Through careful implementation and monitoring, coupled with supported analysis of test results, organizations can glean valuable insights into the impact of proposed changes.

Crucially, the implementation process is not static but iterative, with each test serving as a learning opportunity to inform future initiatives. By embracing a culture of experimentation and continual improvement, retailers can navigate the complexities of the retail landscape with agility and confidence, driving sustainable growth and innovation.

About MarketDial

MarketDial delivers decision intelligence by automating the data science behind in-store A/B testing. Accurate test results measure ROI, enabling retailers to know on a small scale what changes will have big-scale impacts. For retailers asking, "What if," MarketDial provides answers.

Glossary of terms

Capacity planning team (CPT): a team that approves test stores, timelines, and implementation plans for every change that takes place in stores, whether that change be testing or rollout.

Decision committee: a group that helps create and implement consistent decision-making standards across an organization and who approves and oversees testing on change initiatives with high capital expenditures.

Effort prioritization factoring: considering the amount of effort a test will require when determining if that test should be a high or low priority.

Ideator: The individual who comes up with test idea.

Importance prioritization factoring: Considering the level of importance of each change initiative when determining which changes should be tested first.

Project manager: The individual who plans and manages test implementation and tracking.

Store team: The group that implements the operational aspects of the test in stores

Subject Matter Expert (SME): The individual who designs and analyzes the test and builds the business case for project rollout.

Testing interest group: A cross-functional team of leadership and/or users that meets on a monthly-to-quarterly basis to discuss testing topics.

Testing steering committee a cross-functional team of leadership and/or users that meets on a monthly-to-quarterly basis to coordinate testing schedules and avoid interference.