The Optimized A/B Testing Playbook

MarketDial

When companies don't test their brick-and-mortar initiatives before rolling them out across their fleet, they risk making expensive mistakes.

"Many times organizations think they can short-circuit ideas straight to implementation — when testing is the way to make sure that you're implementing something that's not going to break," said Johnny Stoddard, chief data scientist and co-founder of MarketDial. "It can be tempting to think of testing as something that slows you down because it's another step in the process, and it seems expensive. But actually not testing is what can be costly."

Now is the time for companies to embrace disciplined experimentation as part of doing business. To achieve reliable test results, every test should follow a distinct playbook.

Trusting Testing and Data Over Instinct

Large enterprises are run by talented people with no shortage of ideas for improving how business is done.

"Big organizations are generally really good at generating ideas or hypotheses," Stoddard said. "There are reasons why these people are leaders in their organizations; they have novel and interesting ways of thinking about their businesses and thinking up ideas that their competitors don't have."

But while companies are good at asking "what if?" and putting new ideas into action, they can't know if those ideas will succeed unless they test first.

Implementing initiatives on instinct alone — whether they're pricing or merchandising changes, layout redesigns, or other partnerships, or investments — can lead to unfavorable outcomes.

"A famous example of this, in the negative sense, came from JCPenney in 2012, when it reworked all its pricing to be fair and square — so they were no longer going to do coupons or sales or promotions; they were just going to be transparent in pricing." said Morgan Davis, CEO and co-founder of MarketDial. "Sales fell 25% in 2012 compared to 2011, all out of an overconfidence that the company knew what it thought customers wanted and didn't experiment."
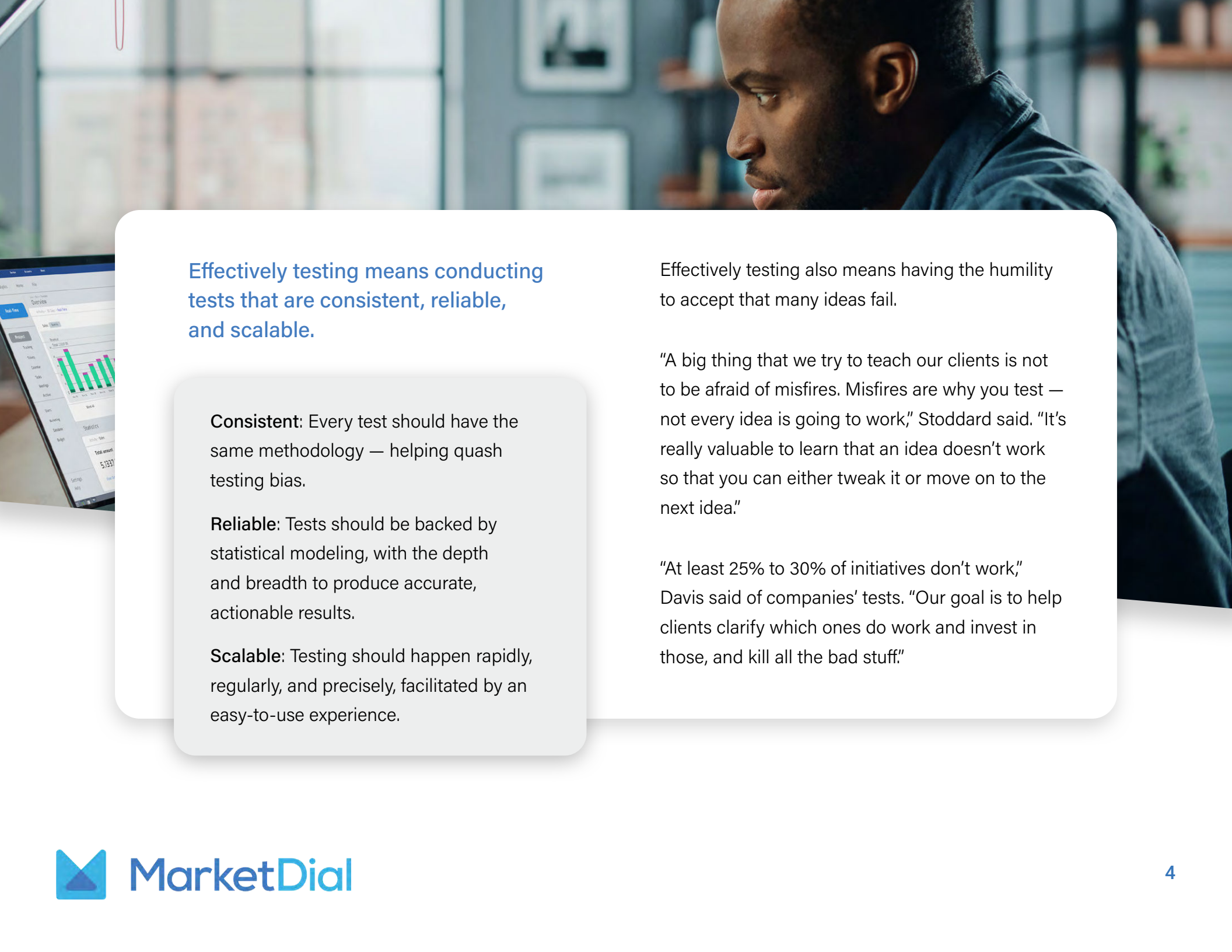
MarketDial

Enterprise A/B Testing Spectrum

enterprises are effectively testing

enterprises fall into that middle-zone group

enterprises fall into the group that isn't testing

Ultimately, many organizations are unaccustomed to trusting testing and data instead of their instincts.

"There's a spectrum that organizations are on when it comes to testing," Davis said. "At the lowest end, some companies — especially newcomers — make a lot of gut-driven decisions and a lot of product-driven decisions, or focus on brand-driven retailing where they don't really care about experimenting. Then there's a middle zone of organizations who think they are experimenting but actually aren't. They're running 'tests' in two or three locations near headquarters; in reality, their tests need to be a lot more rigorous.

"Around 65% of organizations fall into that middle-zone group," Davis said. "Roughly another 30% are in the group that isn't testing. This leaves only 5% that are effectively testing."

MarketDial

Effectively testing means conducting tests that are consistent, reliable, and scalable.

Consistent: Every test should have the same methodology — helping quash testing bias.

Reliable: Tests should be backed by statistical modeling, with the depth and breadth to produce accurate, actionable results.

Scalable: Testing should happen rapidly, regularly, and precisely, facilitated by an easy-to-use experience.

Effectively testing also means having the humility to accept that many ideas fail.

"A big thing that we try to teach our clients is not to be afraid of misfires. Misfires are why you test — not every idea is going to work," Stoddard said. "It's really valuable to learn that an idea doesn't work so that you can either tweak it or move on to the next idea."

"At least 25% to 30% of initiatives don't work," Davis said of companies' tests. "Our goal is to help clients clarify which ones do work and invest in those, and kill all the bad stuff."

A Six-Step Process

01 **Develop Your Hypothesis**

Your hypothesis should define the intent of your test in a clear, measurable way.

"A bad hypothesis is one that is vague, has no numbers, and isn't well communicated throughout the organization," Davis said. "So a good version looks like, 'I am going to raise the price on all king-size candy bars from $1 to $1.50. I am expecting to lose 10% of the units, but to gain 30% in margin.' That is really clear, and it can be clearly articulated to everbody that's involved in the experiment.

"The reason why that's a good hypothesis is that, at the end of that experiment, you can look backward and say, 'Okay, we lost 12% of sales and only got a 25% increase in margin,'" Davis added. "You can know that it didn't meet the target — it didn't meet the goal. And then you can talk about what to do after that, but with the understanding that you didn't succeed in that hypothesis."

MarketDial

02 Create Test Segments

To test your hypothesis, you'll need to first firm up the variables you're testing to achieve the desired hypothesis. This could be increasing the price of a product or entire category to achieve a desired lift (such as with the king-size candy bars) or implementing promotions to increase foot traffic.

Your test segments should focus on one discrete component of how you do business.

"If you go into one market and change the merchandising and change the marketing and retrain employees and reskin the locations all at once, you don't have a good way of knowing which of those four things works," Stoddard said. "A hypothesis needs to be discrete. 'Will this specific change to this part of the business lead to this result?' as opposed to, 'Will these 20 changes lead to some unknowable result?"

In knowing what result you're after, you should also know what actions you'll take should you achieve it (such as implementing price changes or promotions across your fleet) and what actions you'll take should you not achieve it (such as abandoning remodeling plans that fail to create desired revenue changes).
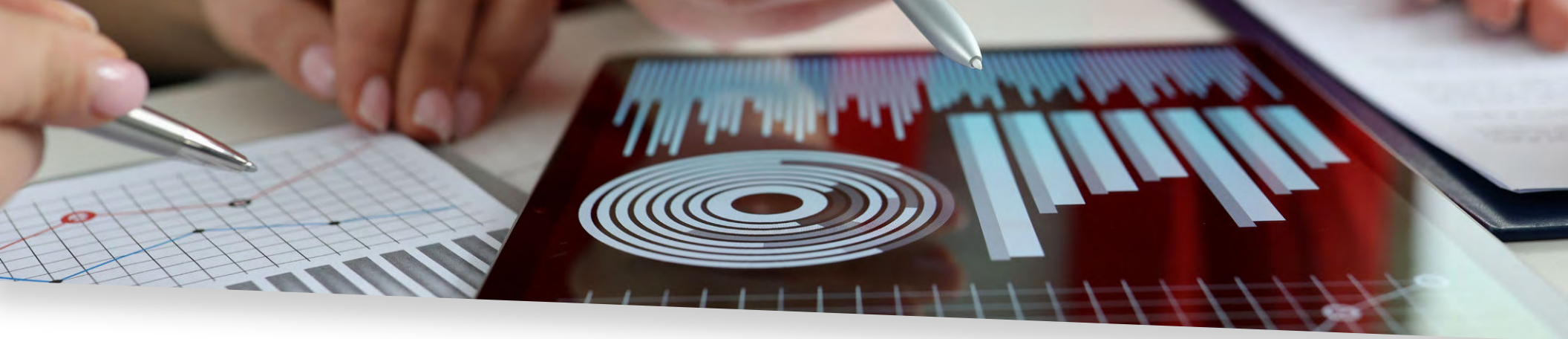
MarketDial

03 Set Test Parameters

From there, you need to determine the specific parameters necessary to achieve your desired result and confidence level. Software can aid in defining these.

"The MarketDial software platform helps our customers do two big buckets of tasks, of which creating parameters is part of the first bucket," Davis said.

"The two big buckets are really designing the experiment overall and then analyzing that experiment." Davis added. "Designing the experiment is a process you go through that involves a bunch of statistical questions, but it starts with 'What variables am I going to be impacting? How many locations do I need to put this in?' and 'How long does the test need to run for?"

"There's a lot that goes into a well-designed test, and sample size and time are key," Stoddard said. "The software helps you determine, 'Do I do this in 20 locations versus a hundred? Can I learn enough by doing a weekend promotion, or do I need to run a promotion for several weeks to reach a certain significance level?'"

04 Create a Representative Treatment Set

The next aspect of testing design is confirming your selection of treatment locations — using criteria that reduce human bias.

"If you're doing a pricing change, and you run it in all your locations that are in highly populous, high-income areas, that's a different result than you're going to get for a pricing change compared with doing it in more rural, lower-income areas," Stoddard said.

"Because the pricing elasticity is so different from those two groups, the pricing change might be absorbed in those high-income areas and not in others," Stoddard added. "So you need a sample that comprehends that and hundreds of other variables — things like customer demographics, how old the location is, and how equipped the location is to implement changes. You need a non-biased sample."

05 Create a Control Match

Since you'll ultimately be comparing results between your treatment locations and a control group of locations, you'll also need the help of software to confirm compatibility between treatment and control sites.

"There's a lot of decisions that go into how you match test locations to control locations," Stoddard said. "We [at MarketDial] solve that through machine learning, advanced analytics, and good data. 'Have these sites been open for a similar period of time? Do they have similar customer demographics? Do they have a similar mix of products that are sold in them? Do the seasonal trends of these locations match each other?'"

MarketDial

06 Execute, Analyze, and Implement Results

With steps one through five completed, you have the test design necessary to execute tests and learn from the results.

"With the test design created, organizations can go out and change the price in those locations or go add labor to some sites, or whatever they're testing," Davis said. "And when that's done, our software analyzes it and determines the incremental benefit, positive or negative, from the price change or extra labor in the location. And that's when we can determine, 'Did it meet the hypothesis goal or not?'"

With that information, organizations can act — by implementing their idea across their fleet or tweaking their hypothesis to continue testing and learning.

MarketDial

In Conclusion

With a six-step, software-powered approach to testing, companies have the opportunity to conduct more frequent and more reliable testing across their organizations.

"At a high level, you need to be able to allocate money based on the results of these tests," Davis said. "And if you don't believe that one team did it right, then you spend all your time arguing about how the test was created and if it's valid. And so having a tool like MarketDial that forces good decisions every step of the way — and that makes sure you're doing it accurately — is really valuable.

You don't need a statistician to enable each test by each team. Each team can build a perfect experiment every time with MarketDial."

Ultimately, executives need to continue asking, "What if?" — and understand that they don't always know the answer — in order to avoid costly mistakes (as consumer preferences can always surprise).

Disciplined experimentation can help organizations make smart choices that drive improvements to their business.

MarketDial