Key drivers for developing into a mature testing organization

Sarah Jacobsen, Data Scientist, MarketDial
Johnny Stoddard, Chief Customer Officer and Cofounder, MarketDial

Introduction

Like developing any skill, the more intentional you are with testing, the more proficient you become. And the more proficient you become, the more you are able to excel. Testing maturity takes basic testing to an expansive level. Note: we did not say an advanced level.

When companies mature their testing capabilities, they expand their reach, their knowledge, their communication, and their decision-making skills across their organization. Testing maturity is less about achieving high-level competency and more about developing ways to systemically collaborate and apply learnings.

Here we establish a five-part framework for evaluating test maturity, explore ways to develop maturity, and provide a self-assessment to help identify growth opportunities.

## Content

The journey to testing maturity

As companies grow into more mature testing organizations, their cohesion and overall testing effectiveness on both a micro and macro level evolves. Maturity matters because it makes testing scalable and repeatable. But development, by definition, is always a process, and no matter where you are in the process, as long as you're growing, you're on the right track.

At MarketDial we've found when companies first begin working with us, they tend to fall into three camps: those who are relatively new to any kind of testing, those who have been testing manually as resources have allowed, and those who test regularly with automated test solutions. While these categories represent a growth in maturity, we've discovered that there is no one-size-fits-all approach to testing. One mature testing organization may look totally different from another in terms of governance, ownership, or types of tests they are running.

We've surveyed companies that have up to 20K sites and those as small as 90 sites, and we've learned that each one has strengths as well as some under-developed testing muscle that can be strengthened. It's akin to human development – learning to crawl, then walk, then run (then ski). The broad range of muscle growth, mental focus, and exercise needed for each skill set is diverse, and the ultimate goal is progression not destination.

Strengthening the testing muscle

The ultimate goal is progression, not destination.

Learn to **crawl**.

Then **walk**.

Then **run**.
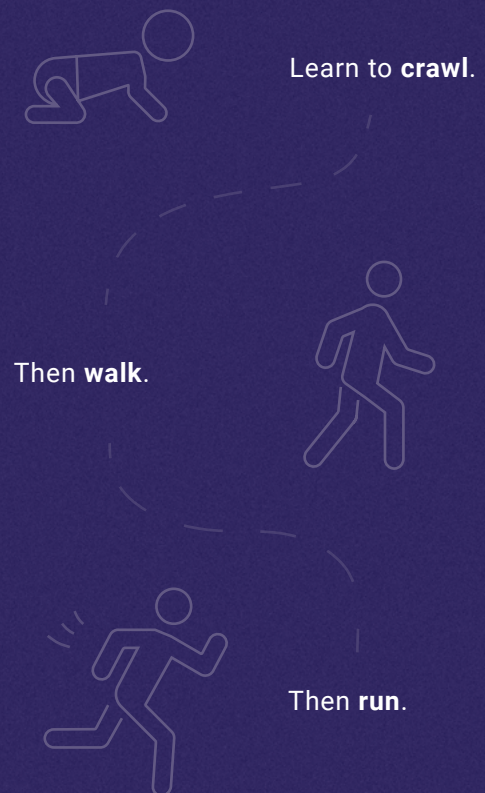
Five main categories for assessing maturity levels

As we have researched both our clients and the broader industry, we've explored what systemic processes have yielded the strongest testing cultures.

What we thought would produce cut-and-dry results in regards to testing maturity instead produced a broad spectrum of growth opportunities for any organization at any place in their testing development.

Based on our research, we believe testing maturity is multi-faceted. Retailers can be more mature in one area and lagging in another. We've established a five-category framework to codify these varying facets and help you assess your level of maturity in each category.

1   Volume and variety
2   Quality
3   Penetration
4   Win rate
5   User sophistication

1 - Volume and variety

Perhaps the most commonly-used metric for evaluating testing maturity, volume and variety measures quantity – the number of tests and the number of departments utilizing testing. The retailers who were most mature had 48 to 278 average active tests per month and had over 60% of their departments testing. Larger mature retailers in this category were building 6 to 23 new tests per month while smaller mature retailers with less than 300 sites were building 3 to 5 tests per month.

We measured volume and variety based on the following criteria:

◇   tests built per month

◇   average number of active tests per month

◇   Percentage of departments testing (marketing, store operations, category management, workforce management, etc.).

Are you crawling, walking, or running?

1-2 new tests per month
0-40% of departments testing

2-5 new tests per month
40-60% of departments testing

5+ new tests per month
60%+ of departments testing

2 - Quality

No doubt you've heard this before: quality matters as much as quantity. How you are testing is as important as how much you are testing. Sample size and test length impact the quality of the results. Consequently, the larger retailers who were most mature in this category had an average sample size of over 53 sites and an average test length of 20+ weeks. Smaller mature retailers had sample sizes of 10 to 20 sites. It's important to note here that even if you don't have many stores, you can still have higher quality tests by testing for longer periods of time.

We measured quality using:

◇   average sample size

◇   average test length in weeks.

Are you crawling, walking, or running?

	average sample size	average test length
	16	10 weeks
	29	17 weeks
	111	23 weeks

3 - Penetration

We've found that as companies become more adept at testing, the number of departments or business functions conducting tests expands, spreading the culture of testing throughout the organization. The most mature retailers in this category had 18 or more total users across departments and over 33% of users actively engaged (analyzing at least one test a month).

We measured penetration by evaluating:

◇   the total number of users

◇   the percent of active users.

On average, mature retailers had 18 total users, with 33% being active users.

4 - Win rate

If every test is predictably positive, then it's likely retailers are not taking enough risks in their ideation processes. If every test is negative, retailers may misunderstand their customer base or be overshooting the mark with innovation. The retailers who were most mature in this category had over 30% of tests as clear wins and up to 15% of tests as clear losses. (The remaining 55% of tests were neutral impact and/or low confidence.) It's also important to note that regardless of lift and confidence, all tests have actionable insights and important learnings.

As one CFO of a C-Store Chain said:

" *I was pumping my fist when a certain test came back with negative results! We have to occasionally see failures to keep us honest. If every test came back positive then we would stop relying on the data in our decision-making.* **"**

We found that small retailers can really nail this category. Since testing volume is lower, smaller companies are often more intentional about what they test in their stores. In fact, we've found that many large retailers struggle with finding good test ideas because they don't filter out enough of the bad ideas. Balance is key when it comes to what to test, how to test it, and what insights are gleaned from it.

We measured win rate with:

◇ the percent of tests run that were clear wins (positive lift with high confidence)

◇ the percent of tests run that were clear losses (negative lift with high confidence).

Average test results for mature retailers.

30%
clear wins

15%
clear losses

55%
neutral results

---

5 - User sophistication

It's one thing to set up and orchestrate a test; it's another to then deep-dive into analytics post test. Savvy testing organizations perform a variety of advanced analytics to dive into second-order questions around test results. They aren't satisfied until they understand what is driving their test results, and they want to scour the data for ideas on what other opportunities exist for creating remarkable results.

We measured user sophistication based on number of users and frequency of use for MarketDial's more advanced features:

◇ Customer Level Testing

◇ Basket Analyzer

◇ Impact Analyzer

◇ Driver Analyzer

Pivotal infrastructure for supporting higher maturity

The most mature companies share commonalities in infrastructure that translate into expanded testing capabilities. We've found that companies who empower their teams, have senior buy-in, and who have distinctive governance all tend to mature their testing at a much higher pace.

Empowering your teams

The more people in your organization who feel empowered to test, the more your testing will excel. And the more your testing excels company-wide, the more employees will want to be engaged in testing.

From your everyday analysts to managers and directors, you can empower your employees by:

◇ ensuring folks are trained in testing practices, methodology, and tools.

◇ allocating budget for testing.

◇ encouraging stakeholder participation.

◇ celebrating wins *and* losses — when people aren't afraid to fail, they have the courage to explore new innovations; losses should be exciting.

Senior mandate

Having support from the top equips those on the ground with the tools required to be effective. Leaders need to be the biggest advocates for testing and must hold their teams accountable. Oftentimes we see those conducting the tests lobbying senior management for more support, using time and energy that could instead be directed towards testing. As a result, those organizations that have senior buy-in from the outset have an advantage over those organizations still clamoring for senior validation and resources. It's also important that leaders take action using test results, showing the company that they value testing enough to make business decisions from it.

Senior leaders can guide testing processes in these meaningful ways:

◇ Gate go/no-go decisions based on testing.

◇ Hold regular status meetings regarding testing progress.

◇ Set an example! Utilize test terminology when discussing initiatives to help others learn the lingo (e.g. lift, confidence, control stores).

◇ Make business decisions using test results. ⏷

Clear governance

Democratizing your data across your company is an essential part of test maturity, primarily because it standardizes test processes and communications, putting everyone on the same page when it comes time for decision making. Having a clear and well-understood governance structure for approving and implementing tests, sharing out knowledge, and coordinating work across teams, will bolster democratization.
In analyzing governance structures among several dozen retailers, we've found there is no universally superior method. *Therefore, we have categorized governance structures into three primary paradigms.*
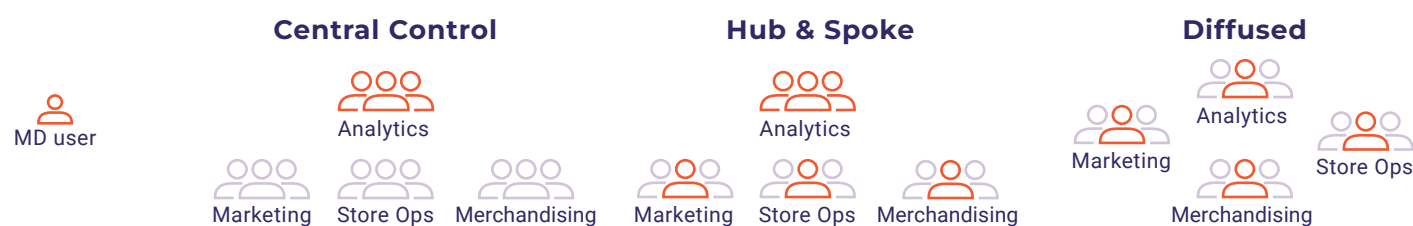
1 Central Control

In a company with a centralized control model, there are typically just a few main users who build 100% of the tests that the company runs. All testing is coordinated by these few users, and they become the solo source for generating test ideas, building tests, and test dissemination.

2 Hub & Spoke

In a company with a hub and spoke model, there is a main group of analysts that oversee testing but ideas and plans are spread across multiple teams. This model can look very different across companies, but an example could be a central analytics team who builds and analyzes tests but who coordinates with members of other teams to brainstorm and implement tests. These other teams may have an experienced user on the team as well.

3 Diffused

In a company with a diffused testing model, there are many users across teams and positions who build and analyze tests. The responsibility of testing falls on every team and there is not necessarily one person or team who oversees or manages all the testing. ⌐

	Central Control	Hub & Spoke	Diffused
Description	Core analytics team field testing requests from functional teams	Core analytics team is COE; designated MD users across teams	MD users across all teams
Pros	Central team ensures consistency and expertise within testing process	Larger testing volume and variety; COE maintains best practices	Highest testing volume and variety
Cons	Low testing volume and less test buy-in from other teams	Requires strong testing governance (Steerco, etc.)	Requires strong testing/ implementation approval process

Why MarketDial?

MarketDial has helped hundreds of retail brands super power their testing. We pride ourselves in challenging retailers to improve their testing volume, variety, quality, win rate, and user experience. Wherever a company finds themselves in terms of testing maturity, we can take them to the next level.

We'd love to hear from you about how testing is going at your organization. Reach out to us, and let's connect on ways we can support you in your testing journey.

More than 100 leading companies and global brands rely on MarketDial to help them create a culture of experimentation with intentionally easy A/B testing. MarketDial's simple interface makes asking "what if?" the cost-effective protocol for intelligent action. Learn how now ⌐

Sarah Jacobsen
Data Scientist

Sarah grew up in Seattle, Washington and graduated with a bachelor's in math education from Brigham Young University and a master's in educational leadership from the University of Utah. After teaching college-level statistics, she knew she had found her passion and is now working on a second master's degree in data science from Eastern University. When she's not studying or playing with numbers, Sarah enjoys spending time with her daughter and creating art through the medium of embroidery. She also loves to read sci fi with a particular affinity for Marvel and Star Wars.

Johnny Stoddard
Chief Customer Officer and Cofounder

Johnny is a data scientist at his core, with a background in management consulting with McKinsey & Company. As a management consultant, he helped executives from retail partners address pressing challenges including: marketing, strategy, commercial transformation, operations, organization, big data, and advanced analytics. Johnny has degrees from Brigham Young University in Economics, Business Strategy, and Visual Arts. When Johnny is not working with data, he can often be found spending time with his children or in the mountains prepping for an upcoming endurance race, mountain climbing, or skiing.

Testing maturity assessment

To see where your company stacks up in terms of test maturity, reach out to us at explore@marketdial.com, and we'll review your assessment.

Volume and variety

◇ How many tests do you run per year?

◇ How many tests are you running per month, on average?

◇ How many departments are testing? List each department along with the number of tests run annual in that department

Quality

For each test you run, rate the following on a scale of 1 to 10, with 10 being the highest quality.

◇ Sample size

◇ Test length

◇ Statistical significance

Penetration

◇ How many people are living and breathing testing in your organization?

◇ How many are active and engaged users of the testing software or tools?

Win rate

◇ How many of your tests produce the results you predicted?

◇ How many of your tests produce positive results?

◇ How many of your tests produce neutral results?

◇ How many of your tests come back with negative results?

◇ Would you say there is a healthy mix of results to ensure you are delivering innovative solutions?

User sophistication

◇ Who in your organization conducts testing and what is their experience level?

◇ How savvy is your organization as a whole at testing?

◇ Do testers ask second-order questions and dig into the root problem?

◇ How advanced are the analytics you run to support your testing program?

Testing maturity assessment

To see where your company stacks up in terms of test maturity, reach out to us at explore@marketdial.com, and we'll review your assessment.

Empowerment

◇ In what ways do you train your employees?

◇ Do your employees feel they have the tools they need to be successful?

◇ How much budget do you allocate for testing?

◇ How involved is each stakeholder?

◇ How do you react when a test fails? How do you support your employees when they face losses?
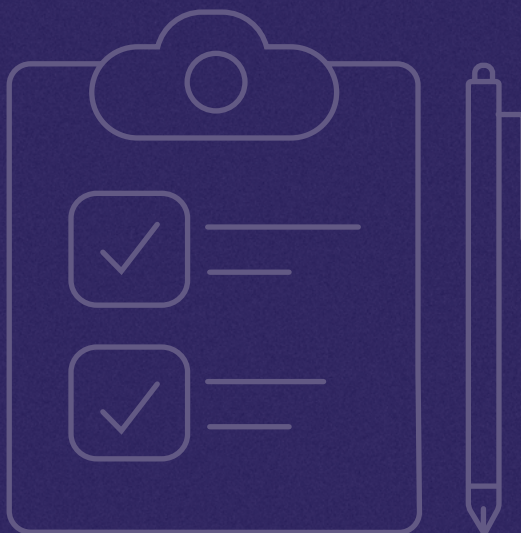
Senior mandate

◇ How do you enable your employees to test their ideas?

◇ How are results communicated and received? Would employees feel a need to hide results of a negative outcome in test?

◇ How do you celebrate negative test outcomes as key learning experiences?

◇ How do you handle regular status meetings to keep employees well informed?

◇ In what ways do you exemplify or model testing protocols?

Clear governance

◇ How would you describe your organizational testing structure?

◇ Which system most aligns with your structure (see Page 7 for description):

Central Control
Hub & Spoke
Diffused

◇ Does your current governance structure suit your testing needs?