

B

The State of Experimentation Report

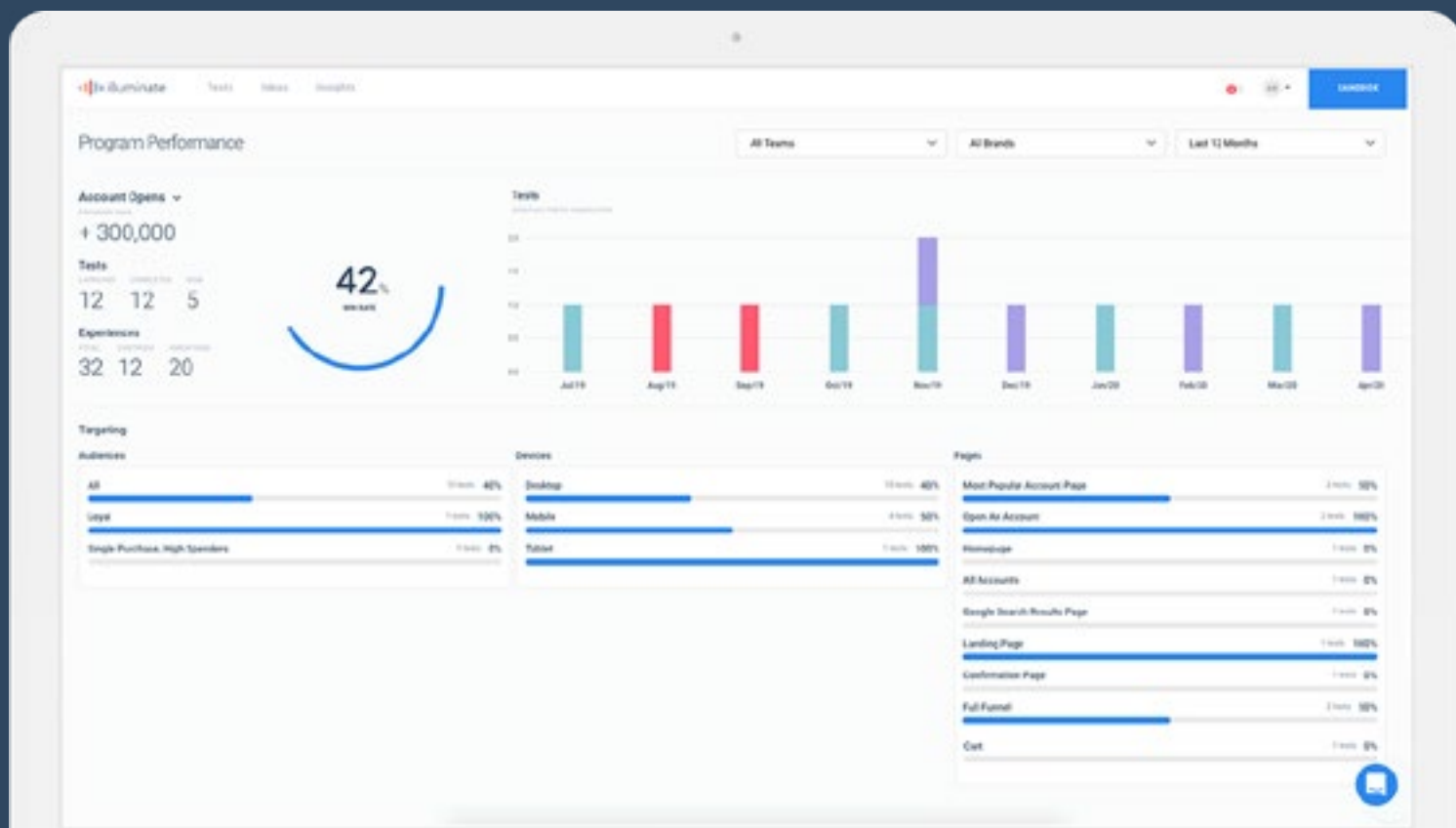
2019 - 2020 Edition

Published July 14, 2020

BROOKS BELL

This report is powered by illuminate

Run a more organized and impactful testing program with illuminate, our **FREE** experimentation program management software.



Introduction

Digital experimentation, or testing, when well-executed, can be like a playbook for your business. While it may not give you all the correct answers, it provides an organized way to make choices with minimal risk and maximal opportunity.

Testing is a powerful tool that can generate consumer insights, answer tough questions, drive progress toward business goals, and ultimately, lead to increases in revenue or other indicators of performance.

Of course, if you're running a successful testing program, you already know this.

You may already have a culture where decisions are driven by data, and tests are meticulously cataloged, aggregated and archived. You may even be benchmarking your progress and calculating your improvement over prior years.

But how does your program compare to others? Other peers, other industries, other testing techniques?

This report offers an opportunity to find out. We looked at a sample of nearly 500 tests, spanning industries,

launched between January 1, 2019 and January 1, 2020.

This sample allows us to establish both broad and industry-level benchmarks. We've also broken down the trends by the strategy used in designing the test and the page on which the test was launched.

Of course, none of this would be possible without data. For this report, we utilized an anonymous sample of data from illuminate®, [our free A/B testing program management software.](#)

We built illuminate to help testing leaders run a more organized and impactful experimentation program. The platform enables you to create a searchable library of past, current and future tests, track your testing activities against program KPIs and company goals, and build, store and quickly share interactive reports, case studies and customer insights.

While illuminate is uniquely suited to the needs of testing teams, an important takeaway is that without a system of record for your testing efforts, a retrospective assessment is simply not possible.

A little context regarding the data in this report:

While the sample spans industries and businesses, there is a selection bias that must be considered.

For the most part, testing teams using illuminate are experienced, organized and achieving an above average—if not exceptionally high—level of operational maturity.

Additionally, the sample mostly includes large to very large businesses. This means that the ratio of win-loss-flat may be different in this report as compared to benchmarks that consider a broader diversity of testing programs, such as those of newer programs or small to medium-sized businesses.

Finally, this analysis pertains to data collected between January 1, 2019 and January 1, 2020, and as such, does not take into account the COVID-19 pandemic and subsequent economic downturn.

All that said, we feel the this analysis provides an upper bound of long-term performance, and an aspirational benchmark for less-mature testing programs.



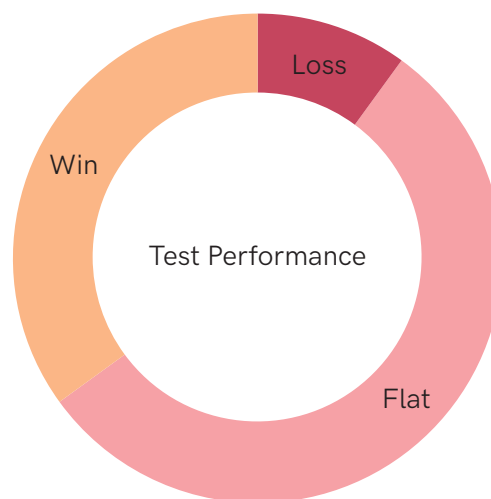
Overall Testing Performance

When we look across all industries in our sample, there is some encouraging news. Overall, tests win far more than they lose. In our sample, 35% of all tests were declared a winner, while only 11% were declared outright losers.

There are a few important lessons in these numbers. First, the oft-cited fear of many managers that testing will harm business performance is not supported by the evidence. Through testing and measurement, we can clearly see there is a much greater chance a test will produce a positive result than a negative one. In fact, based on this data, if we consider the odds of a win versus a loss, we would expect slightly more than four wins for every one loss. That said, an overall win rate of 35% may still seem low. If only 35% of all tests are declared winners and 11% declared losers, what happens to the majority of tests?

The answer is that about 55% of tests in our sample were declared “flat,” or more precisely, a statistical difference between the control and challenger could not be determined in the predefined test period.

While this may seem discouraging for testing teams, we like to think of it like an offensive play in football. Yes, you could go for the Hail Mary pass with greater risk



Data: Brooks Bell's 2020 State of Experimentation Report

and greater reward. But most likely, you're going to incrementally work your way down the field, using safer plays to gain yardage incrementally and testing different strategies on your opponent to find what works.

Further, finding statistical winners is not the only positive outcome of running a test. If carefully designed, every test should produce some insight, regardless of whether the result is declared as win, loss or flat. That said, the goal is to win and guidance as to what differentiates the winning test from a losing or flat test would be helpful. To try to derive some of that insight, we will break down the overall win rates by industry, strategy, and test location.



Testing Performance by Industry

Obviously, these industries have different goals, challenges, and consumer decision journeys. The industry baseline risk tolerance varies, as does the flexibility for testing and implementing successful strategies. In addition, businesses in different industries operate on different cycles, meaning the volume of tests may differ substantially. Indeed, we see this clearly in the data.

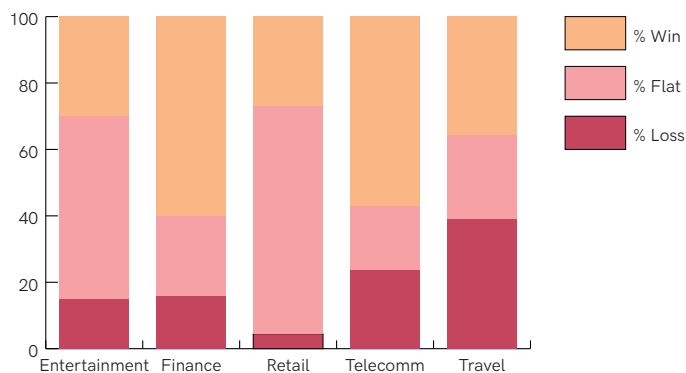
Overall, retail executes the most tests, representing 61% of our total sample. Next is finance, which represents 16% of the sample, followed by entertainment representing 13%, travel representing 6%, and telecommunications representing 4%. Volume, however, does not seem to

directly predict success (see Figure 1), which is clear if we compare the percentage of wins in retail (27%) to the percentage of wins in, say, finance (60%).

There are a number of reasons why the win-loss-flat rates vary so significantly across industries. One helpful theory is based on the nature of the consumer decision journey. For retail, for example, consumers arrive at the storefront (website) earlier in the decision process. Perhaps they are beginning the information gathering phase or, more likely, they have not yet clearly identified a need. When making a retail purchase, as compared to a financial conversion, the distance between “browsing” and “buying” is orders of magnitude shorter. Thus, retailers benefit by testing often, knowing on the one hand that minor changes will make a large difference in purchase behavior, but on the other that finding the changes that will have an impact will be very difficult. The upside for retailers is that it is similarly difficult to negatively impact purchase behavior, as demonstrated by the very low loss rate for that industry (4.5%).

Another possible explanation for the difference in win-loss-flat rates is that the focus of testing varies across industry. To better understand that, we need to consider the proportion of pain points addressed across the five industries in our sample.

Figure 1. Win-loss-flat rate and testing volume by industry.



Data: Brooks Bell's 2020 State of Experimentation Report



Testing Focus by Industry

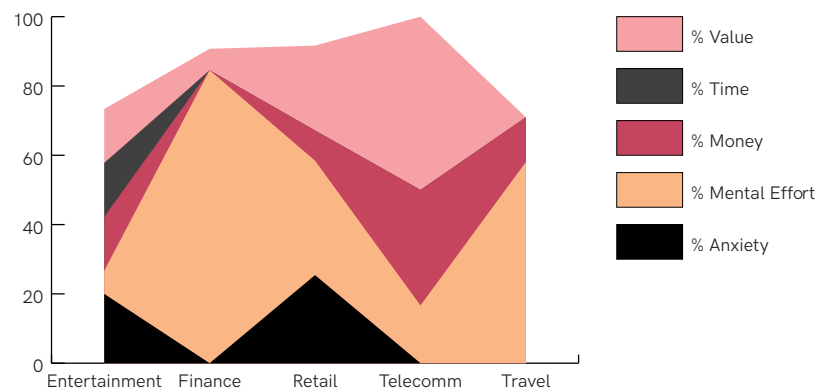
At Brooks Bell, we have defined a multi-layered taxonomy of testing strategies, which informs the assessment of consumer experiences, decision problems, test ideation, hypothesis creation, design and development, and more. It standardizes our approach to testing. It also provides a useful framework for comparing test strategies over time. For the purposes of this report, we have analyzed test performance at the highest level of the taxonomy: the consumer pain point the test was designed to address.

These pain points are anxiety, mental effort, money, time, and value.

Anxiety includes fears and uncertainties a consumer may have about the product, purchase, or decision.

Mental effort includes everything that may make the decision more

Figure 2. Pain points addressed as a proportion of total tests by industry.



Data: Brooks Bell's 2020 State of Experimentation Report

Note: Values may not sum to 100 due to sample masking to preserve anonymity and other factors.

cognitively difficult, including having to compare information across pages, text that is difficult to read or understand, and convoluted checkout procedures.

Money relates directly to the transactional component of a decision, most often the price or cost.

Time refers to both the time-as-cost, as in the time it takes to complete

an action and time-as-value, which happens when a decision promises to save time in the future.

Value refers to the benefit or benefits received as a result of making the purchase or decision.

When we look across industries (see Figure 2), a few trends become clear. Perhaps most obvious is that in many industries, tests tend to focus on only a few possible consumer pain points.

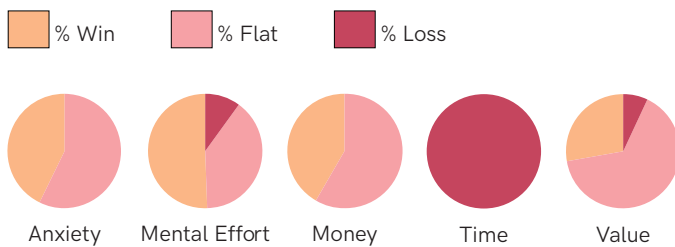


A few takeaways:

- Time, either as a cost or value, is underutilized.
- Finance focuses exclusively on mental effort and value.
- Travel focuses exclusively on mental effort and money.
- Mental effort represents the vast majority of test ideas. This is perhaps not surprising, since mental effort typically includes most changes to design elements and site functionality. While messaging and other creative changes can also fall within mental effort (e.g., making the call to action more clear), strategies addressing the other pain points almost exclusively require a messaging or creative change.

The focus on mental effort in our sample is likely not a simple byproduct of the nature of the test designs it includes.

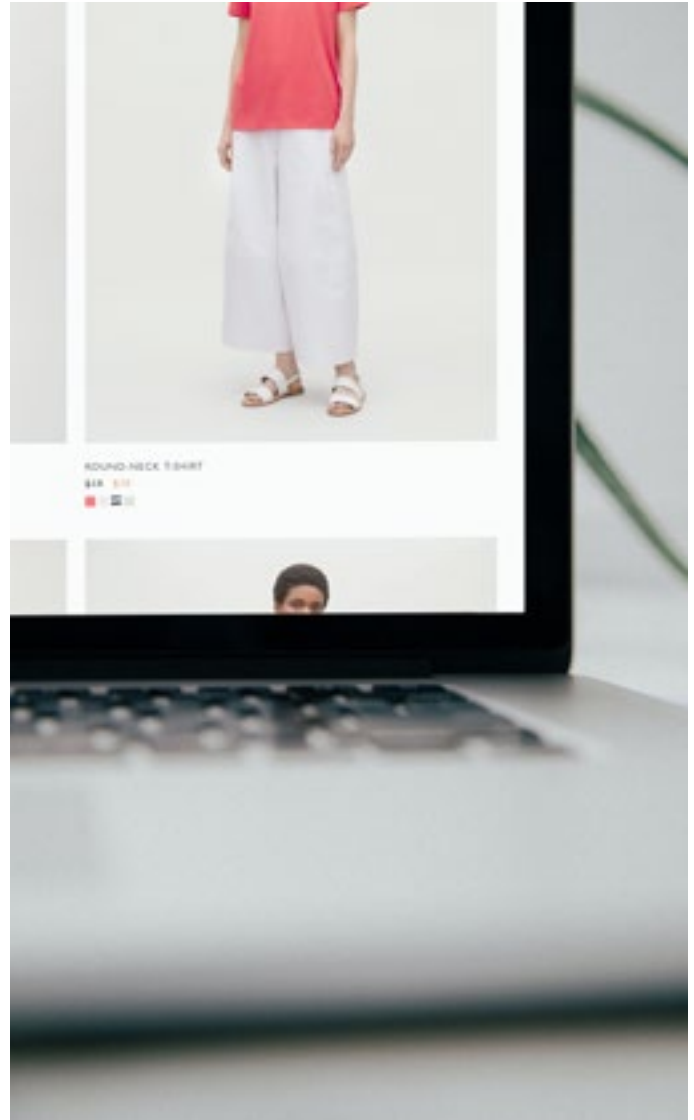
Figure 3.
Win-loss-flat rate by pain point.



Data: Brooks Bell's 2020 State of Experimentation Report

Indeed, when the win-loss-flat rate is considered by pain point, we see that mental effort produces the most wins (50%; see Figure 3).

Moreover, the risk is in line with the overall loss rate (10%) and the incidence of flat results is below the overall flat rate (40%).



However, not all wins are the same.

Many statistical wins occur near the top of the funnel and are calculated based on engagement KPIs, not sales conversions, for example. Others, given very large samples, represent only a few fractions of a percent of improvement.

Teasing out the dollar value of a test requires carefully guarded information. However, we can get a sense of the overall impact by considering the page the test was implemented on.





Testing Performance by Page Type

One of the most important decisions in designing a winning test is choosing the page that will be the focus. There are several reasons for this.

First, different pages on a website receive substantially different amounts of traffic. The homepage or a landing page tied to a paid-search campaign may receive thousands or millions of views per day. The bottom half of a product support page may only receive a few dozen views per year.

Second, different pages of the site represent literal and symbolic levels of the purchase funnel. Visitors to a landing page tied to a paid-search campaign may have low awareness of the business and little or no engagement with the purchase decision when they arrive. Visitors to the homepage likely have brand awareness and at a minimum curiosity in a purchase, if not an identified need. Visitors in the cart or checkout are signaling some intent to complete a purchase.

Third, behavior is malleable to different degrees on different pages and at different stages in the funnel.

8 Types of Webpages

Site-wide refers to tests that influence some element across the entire website, for example the main navigation or load-time optimizations.

Homepage is the highest-level page in the primary site taxonomy.

Category pages serve as landing pages for a specific class of products or services

Landing page is the highest-level page in a linked but adjacent taxonomy, often the target of paid-search or other advertising campaigns.

PDP is the product description page, or the page featuring a single product or service.

Cart is the shopping bag or shopping cart page in which selected products are accumulated before purchase.

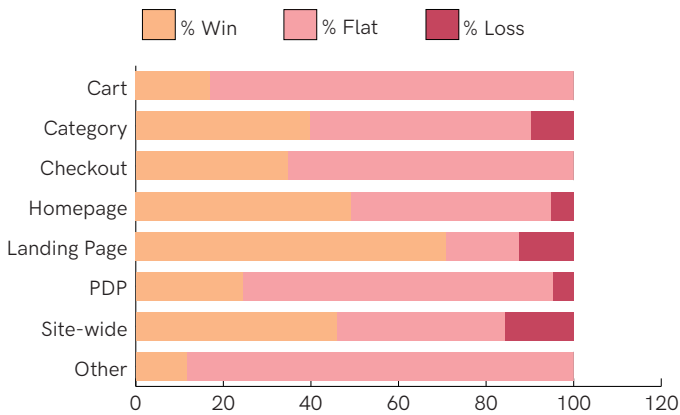
Checkout is the actual payment funnel during which a purchase is completed.

Other includes a handful of other common pages including calendar pages.

Increasing basket size in the checkout, for example, is considerably more difficult than on a product page.

Looking at the total test volume per page (see Figure 4), most tests, by a small margin, occur across entire sites. This could be a sign that businesses are looking to optimize universal site headers, navigation, promotion

Figure 4.
Win-loss-flat rates and testing volume by page.



Data: Brooks Bell's 2020 State of Experimentation Report

boxes, and other elements that impact the site as a whole.

Apart from this, we see that the most tested individual page is the product detail page (PDP). This makes sense since it is often the focus of the consumer decision, especially for retail. Further, the relative complexity of product pages offers a wealth of opportunities for testing.

Homepage testing and category page testing also appear to be relatively common. Notably, the homepage offers a relatively high proportion of wins (49%) to losses (5%), making it a surprisingly low-risk, high-reward location for testing.

Conclusion

Testing is a powerful tool that can fundamentally change business strategy and operations. However, simply running tests does not guarantee success. Even the most successful programs will see a series of wins, losses and inconclusive flat results over time.

Making sense of these results requires aggregation and, ultimately, comparison to others. By considering the benchmarks presented in this report, you can gauge the performance of your own testing program as compared

to others in your industry and to testing programs overall. What's more, by looking at how various strategies have been applied, and where, ideas for new or more targeted test ideas may emerge.

Whether you're beating the benchmark or striving to meet it, better performance can only come one test at a time.

The Five Pains & Gains

Mental Effort

Are users easily able to determine where to go next?

For example:

- Users on a homepage are likely looking for direction. Does your experience provide the right cues?
- Users purchasing routine items may desire a simple, quick experience. Does your experience reflect the purchasing style of your users?

Alleviate the mental effort pain point by:

- Reorder information on the page to place the most important elements above the fold
- Reduce the number of options available as a next action
- Emphasize the next step by redesigning the CTA

Money

Are your users price sensitive? Does converting require a large investment?

For example:

- Users that are very price conscious may be looking for savings. Will they easily see them on your experience?
- Higher price tags may have your users interested in the details. Does your experience go the extra mile to validate their purchase decision?

Alleviate the money pain point by:

- Reformat discount presentations to make them easier to understand
- Adjust the size (or perceived size) of a discount
- Make the price more (or less) prominent

Value

What is it about your brand that makes it the right choice?

For example:

- Users may be considering your experience because your brand is viewed as the expert. Does your experience convey that expertise?
- Users may be bouncing in between sites, unsure of where to convert. Why should they convert with you?

Reduce concerns about value by:

- Add detailed product images that emphasize quality
- Emphasize the functional benefits of a product or services
- Create a comparison that demonstrates the benefits of a product or service

Anxiety

Is it a high consideration purchase or decision, causing increased anxiety about making the right choice?

For example:

- Users making important life decisions on your experience may be looking for reassurances. Do you provide them?
- Stolen data and identity theft has left some users full of worry when converting online. Does your experience build their confidence?

Alleviate the anxiety pain point by:

- Reduce the threat of regret by emphasize return policies
- Build confidence by ensuring design consistency across decision points
- Emphasize the satisfaction of past customers to provide social proof

Time

How does your offering help people spend their time doing things they enjoy?

For example:

- Users may dread a long application process. Does your experience help them prepare for the time commitment required, and remove unnecessary obstacles?
- Your product may replace a time consuming task, allowing them to spend more time with their loved ones. Do you make it easy for them to see those benefits?

Reduce concerns about time by:

- Provide clear cues as to the time a process will take
- Minimize the length and complexity of checkouts and forms
- Emphasize the potential time savings a product or service represents





Transforming businesses through experimentation and insights.

www.brooksbell.com