Experimentation

How Managers Can Build a Culture of Experimentation

by Frank V. Cespedes and Neil Hoyne

February 15, 2022

MirageC/Getty Images

Summary. Companies tend to allocate testing time and money to big initiatives while ignoring small ideas that, in the aggregate, can have a bigger impact with less risk. The hardest challenges in testing are internal processes, especially the need to work collaboratively... **more**

Most managers are good at asking questions, but not as good at specifying what would constitute a feasible answer to those questions. They run tests and learn from the results, but rarely do the results promote ongoing dialogue and organizational change.

Experimentation encourages innovation, but it can also be time and resource-draining. To make experimentation a productive activity in your organization, you must manage several conditions. What you learn from experiments, how you apply your learnings, the opportunities presented by your learnings, and, perhaps more importantly, the conversations you have with colleagues about your findings should have an impact on organizational decision making.

In this article we detail the conditions under which experiments should be conducted, including the key considerations one must make when testing in an ever-changing business environment, which data to test against, and the criteria against which one should make decisions.

Understand testing in business conditions.

Many managers assume a test of a new product, price, or service is analogous to a clinical trial in medicine where a hypothesis can be rigorously validated. But testing in business presents qualitatively different challenges than those in most academic and medical research. There are few opportunities for randomized control trials in a competitive market. You must typically repair the ship while it's sailing on open waters in weather conditions that you do not control. This is especially true in an era of big data and artificial intelligence.

Some tests may be unnecessary or of minimal managerial impact if you consult existing data and literature about the topic. In developing its digital strategy, a prominent retailer did not consider existing research, including a published and peer-reviewed study on cross-channel consumer behavior across more than 7 million purchases by nearly 1 million customers. Insisting that all evidence be "first-hand" data, it commissioned a test using six company-owned locations. Not only did this extend decisions and actions by eight months, without a gain in evidence, it also provided ample opportunity to reinforce legacy biases while competitors began their multi-channel initiatives, impeding this retailer's growth.

You should consider the opportunity costs inherent in testing and be willing to adjust methodology and scope accordingly. A B2B SaaS company was presented with evidence that a traditionally unprofitable customer segment was starting to shift its purchasing behavior, and a

relatively modest marketing investment could accelerate that shift. But the legacy of losses loomed large, so decision-makers set a high bar in terms of experiment duration, sample size, and methodology to overcome organizational disbelief, when much simpler means were available to test the ROI of new initiatives in this segment. The larger tests cost nearly five times more and delayed action in a fast-changing market.

Managerially actionable testing will rarely have a "scientific" result, but it can still yield insights and options. The goal is to generate relevant dialogue among decision-makers in changing market conditions, not eternal truths. Use what you can for today while investing in finding answers for tomorrow.

Mind your data.

You need reliable data to avoid the garbage-in-garbage-out syndrome. In most machine learning projects, as much as 80% of the time and costs of data scientists and IT groups is spent cleaning up the data, due to things like inconsistent inputs, outdated views of buyer behavior, and legacy assumptions.

Common examples involve tests driven by data in customer relationship management (CRM) systems. The inputs are noisy because the system reports the aggregate result of what, in reality, is multiple people using different criteria. One rep inputs a request for a price quote as a qualified lead or active account in the system; another uses identifying a budget as the criterion for qualifying a lead or responding to price queries. The problem is magnified when a multichannel effort is relevant to the test.

Most CRM software also weights revenue expectations by pipeline stage on the assumption that the odds of closing increase in reported

successive stages. But rather than moving sequentially through a linear funnel, omni-channel buyers now move from online to physical to influencer channels multiple times in buying journeys. Once the system is in place, however, tests are then designed to optimize for the software parameters, reinforcing an outdated view of consumer behavior. The test becomes a self-fulfilling prophecy, not a window on market realities. More generally, as others have noted, as easy access via mobile devices makes just-in-time information a growing factor in purchase decisions, many traditional research techniques like conjoint analysis do not reflect how buying decisions are made.

The legacy of such "tests" can linger for years. One of the authors of this article worked with a company where the churn rate in its customer base (3% each year, its marketers alleged) was established back in the 1990s — a figure the company has used ever since, despite repeated changes in products, prices, competition, substitutes, and consumer choices.

You must build tests from data in which you're confident. For example, product returns are a trillion-dollar issue for retailers worldwide — and getting bigger as Amazon Prime makes "free" returns a growing norm. You can ask customers if they plan to return their purchase, but ex-ante surveys are a poor basis for predicting this behavior, and some companies now offer discounts to customers who give up their right to return a product — an inhibition to buying in many categories. A buyer's order history is a firmer basis for testing. One study found that when shoppers interact with products, zooming in to see the texture of the fabric or rotating it to see its appearance from multiple sides, they are less likely to return the purchase. Conversely, those who order in a scattering of sizes are more likely to return products. This data can provide hypotheses for relevant tests that, in turn, generate dialogue

about website design, pricing, order-fulfillment policies, and terms and conditions.

Establish decision-making criteria.

In our experience, the hardest challenges in testing are internal processes, especially the need to work collaboratively to define a problem. Most managers are good at asking questions, but not as good (or, for various reasons, reluctant) at specifying what would constitute a feasible answer to those questions.

Data, even allegedly self-correcting data as in some AI programs, is never the same as the answer to a management issue. Years ago, Peter Drucker emphasized this: "The computer makes no decisions; it's a total moron, and therein lies its strength. It forces us to think, to set the criteria." Data is crucial, but it's mute. Managers must always interpret data with an end in mind.

Pricing is an example. A price has multiple dimensions: base price, discounts off list price, rebates tied to volume, special offers, price for additional services, willingness to pay depending upon the product application, and so on. Further, price information is now often a click away for customers. Sites including Edmunds.com, and Kayak facilitate price comparisons in multiple categories. And inertia is rarely the profit-maximizing option for sellers. Notice, for instance, how Amazon distills thousands of SKUs for consumer-packaged goods into price-per-ounce comparisons on its website.

Price testing should be an ongoing part of effective marketing, but first clarify the evaluation criteria because testing in business ultimately means evaluating alternatives. There's a big difference between using profit increase or revenue lift, for instance, as the criterion, and price

changes typically have an impact over multiple time periods, not just in the short term. Yet, most companies fail to specify the criteria they will use to interpret pricing tests and they spend time and money in an unfocused fishing expedition that goes nowhere.

An exception is Basecamp, the collaborative software provider whose products span a wide range of users, applications, individuals, and large corporations. When it introduced its Basecamp 3 product, it conducted a combination of price surveys, A/B tests, various offers, and specified its criteria up-front for making decisions. As its chief data analyst noted in a Harvard Business School case, Basecamp's products are sold via a low-cost inbound e-commerce model, so "optimal prices [are] those that result in maximum [customer] lifetime value (LTV). We'd accept a lower purchase rate if a higher average value offset that, and vice versa. We'd also accept a lower average invoice amount if it led to higher retention and thus greater LTV." The firm was also clear about the criteria to use in evaluating results: "It's hard to test LTV directly [because] that's a long-term outcome sensitive to elements beyond price... Impact on LTV is estimated by evaluating conversion rates (free accounts who upgrade to a paid plan) and initial monthly revenue (average price a user pays after conversion to a paid plan)."

These criteria helped the organizational dialogue and improved cross-functional efforts to evaluate the data and implement options. There's a tradeoff between LTV pricing opportunities and maximizing initial customer acquisition. Different functions (sales, marketing, operations, finance, investor relations) usually have different views of that trade-off, and in many firms valuable options are stopped by managers who optimize their function's metrics, not enterprise value.

Pay attention to "small" ideas.

Few billion-dollar opportunities start that way, but companies tend to allocate testing time and money to big initiatives while ignoring small ideas that, in the aggregate, can have a bigger impact with less risk. Pricing is again an example. The impact varies by industry, but studies indicate that for a global 1000 firm, a 1% boost in price realization — not necessarily by increasing price on every order, but averaging out to 1% more and holding volume steady — typically means an 8% to 12% gain in operating profits. These results have been consistent for decades — before the internet became a commercial medium, since then, and for both online and offline firms.

Seek progress, not perfection, and invest in processes that allow employees to submit seemingly small ideas. Online channels make testing these ideas feasible and inexpensive when you know how to ask questions. Here are three straightforward approaches:

- Mine your website purchase interactions. When airlines add a question asking if a trip is for business or personal, they have insight into price sensitivity for upgrades.
- Rotate periodically the questions you ask, gathering insights that are missed when the same questions are unchanged for months or years.
- Engage users and non-users. There's now a class of tools that enable you to engage directly with customers and prospects in real time and at different points in their buying journeys.

As the pandemic demonstrated, markets move faster than ever and it's your job to adapt. Talk about "big data" and "digital transformation" has many managers obsessing about how to store data. But the best firms obsess over how they can *use* their data in actionable tests of new ideas. Think of testing in your organization as part of an ongoing conversation
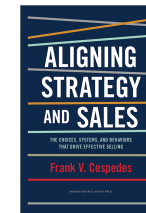
with your market — a motion picture, not a selfie or snapshot, in a world that never stops changing.

Frank V. Cespedes is a senior lecturer at Harvard Business School and the author of _Sales Management That Works: How to Sell in a World That Never Stops Changing_ (Harvard Business Review Press, 2021).

Neil Hoyne is Chief Measurement Strategist at Google and the author of _Converted: The Data-Driven Way to Win Customers' Hearts_ (Penguin).
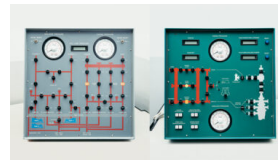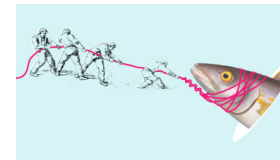
Recommended for You

When Subtraction Adds Value

So Your Boss Offered You a Meaningless Promotion

Drive Innovation with Better Decision-Making

The Two Traits of the Best Problem-Solving Teams

Partner Center

Start my subscription!

Explore HBR
The Latest
Most Popular
All Topics
Magazine Archive
The Big Idea
Reading Lists
Case Selections
Video
Podcasts
Webinars
Data & Visuals
My Library
Newsletters
HBR Press
HBR Ascend

HBR Store
Article Reprints
Books
Cases
Collections
Magazine Issues
HBR Guide Series
HBR 20-Minute Managers
HBR Emotional Intelligence Series
HBR Must Reads
Tools

About HBR
Contact Us
Advertise with Us
Information for Booksellers/Retailers
Masthead
Global Editions
Media Inquiries
Guidelines for Authors
HBR Analytic Services
Copyright Permissions

Manage My Account
My Library
Topic Feeds
Orders
Account Settings
Email Preferences
Account FAQ
Help Center
Contact Customer Service

Follow HBR
Facebook
Twitter
LinkedIn
Instagram
Your Newsreader