A measured approach to retail growth: Why retailers should implement a culture of scientific measurement

By Kim Frazier, Michael Greene, Anna Hong, and Drew Gaputis
Retailers have long sought to improve sales by tinkering with all aspects of layout, product selection, and price. With regards to pricing, effective strategies seem to be an elusive, constantly moving target. Headlines are continually peppered with retailers announcing new—often reactive—pricing, aimed to capture customers and fend off competitors. At times these changes are nuanced, as we see in the retail industry, where companies (apparel retailers, for example) seek to simply refine their high-low, promotional-based approach. In other cases, we see larger strategic shifts—such as the move to everyday low pricing (EDLP) among many general merchandise, grocery, or home improvement retailers. Then, of course, there are massive strategic realignments, some of which have resulted in turmoil and disaster—as we have seen in recent years.

These changes have only accelerated in the age of omnichannel retail and price transparency. Now, pricing strategies should become even more creative and differentiated, as blunt tools such as the “price match” have an uncertain history of success. In light of this constant change, how should retailers decide what path to pursue? Scientific approaches to understanding customers, such as Paco Underhill’s pioneering work in *Why We Buy*, provide a path forward by recommending experimentation and a scientific approach: trying out new ideas to constantly improve. More recently, articles have reinforced the importance of effective experimentation in all aspects of business. However, there is still an opportunity to directly apply the principles of testing and measuring in the world of retail pricing. Despite the best intentions, experimentation among retailers is still characterized by a limited scope or a narrow view of success. For instance, small scale periodic testing may fail to shed insight on the breadth of opportunities available. Elsewhere, scorecards and KPIs rely on approaches such as year-over-year or “same store sales,” which can misrepresent the real impact of a test when exogenous changes such as the economy come into play.

Whether seeking to refine their current course, completely shift strategies, or introduce creative pricing offerings, we recommend retailers focus on establishing a culture of scientific measurement, allowing them to “test their way forward” to pricing success. This goes beyond running a single test. The organizations we have seen succeed have committed to testing multiple opportunities and rigorously measuring success through a “test” and “control” methodology.

Pricing challenges and intense competition in retail markets have only increased in the emerging age of omnichannel retail and price transparency. To be successful in these conditions, pricing strategy needs to become more creative, differentiated, and based in a scientific test and measurement approach.
What we mean by rigorous measurement
Some retailers feel they already measure by tracking common metrics like same store sales. The unfortunate truth is that these metrics are not built to accurately assess the impact of a pilot price test. Broad measures such as year-over-year sales often fail to account for external or unrelated factors that may have influenced sales during the pilot.

To accurately measure the impact of a pilot pricing program, we recommend utilizing a test and control methodology. Test and control is used in scientific experiments to find out the answer to the “what if” question. In Back to the Future, Marty McFly travels back in time and finds out (spoiler alert!) what would happen if his parents had never met. We can apply the same methodology in retail pricing to ask the question, “what if?” To know the impact of a pilot program, we compare what happened in the “test” stores vis-à-vis the “control” stores where the pilot was not in effect. For instance, if a store had a promotion during a given week, how much more did it sell than a store without the promotion?

Retailers can learn from peers in other industries. Internet and technology companies commonly use “A/B testing” where visitors to a website see different pages (note: this methodology is also used among online retailers). The technology company uses differences in behavior to improve user experience and click-through rates. Financial institutions often employ “champion-challenger” schemes, where the newest innovation is tested on a randomly selected cohort to evaluate the program. One of the most rigorous applications is in the pharmaceutical industry, which has to meet the stringent requirements of the FDA via test and control.

While not a new measurement concept, many retailers still face challenges in establishing robust test and control practices.

If it’s so impactful, why don’t all retailers use test and control?
Conducting tests in retail is hard. There are costs in time and dollars. In addition, there may be operational issues to overcome when attempting test and control.

1. Is it worth the upfront costs? One advantage of establishing a broad culture of measurement is the impact it can have in business results. Once a retailer has firmly established a culture of measurement, it can be leveraged for dozens of applications, ranging from pricing, discounts, human resources, training, distribution, and other operational areas. This aspect of test and control allows retailers to spread upfront costs such as acquiring knowledge and resources over a wide range of potentially successful initiatives, not to mention the opportunity costs of avoiding ineffective programs. Establishing a repeatable process can ultimately enable an organization to test a larger number of initiatives in a shorter amount of time, identifying where to focus resources.

Once a retailer has firmly established a culture of measurement, it can be leveraged for dozens of applications, ranging from pricing, discounts, human resources, training, distribution, and other operational areas.
2. Tests must be realistic to be useful. The next issue facing nearly all pilot programs is potential impact. Large retailers could have thousands of brick and mortar stores each requiring local considerations. For example, when a retailer with thousands of locations was changing prices, we faced the prospect of having to update price tags on a million bottles of shampoo already on the shelf. To achieve this, printers ran 24 hours a day for weeks to get the labels right. At another retailer, price tests were accompanied by marketing materials, store selling tools, technology job aides, and in-store training. These tools needed to be in place in advance of the test to ensure a seamless selling process.

3. Time is money. Lastly, and possibly the most difficult issue: testing requires patience. Many price tests can be concluded in a few weeks, but we have measured some price tests for as long as 12-18 months. The willingness to be data-driven and not react to trigger reflexes can be a major cultural change for some retailers. Test-level changes may take time to be fully absorbed by customers, and behavior may change slowly, highlighting the need for testing discipline. In the end, adhering to the principle of patience is all part of establishing a true measurement culture.

**Store selection and measurement**

Once a retailer decides to implement test-control processes, the next question is often: “How do we pick test stores vs. control stores to ensure accurate results?”

- **Random selection.** Since the 1940s, scientists have been using random selection for experiments and measurement. For a retailer, this essentially means flipping a coin and assigning each store to test or control based on heads or tails. The only difference is the pilot program given to the treatment group. The two groups of stores are identical in almost every other way, including the external environment such as the economy.

- **Propensity scoring.** More recently, academics have been pioneering ways to select test and control groups when randomization is not feasible. If we could find a twin for each store in the test group, then we would have a comparable baseline as good as if it were selected randomly. In spirit, that means identify a set of control stores as close as possible in every way to the test stores. Propensity score matching is an algorithmic way to derive test and control groups which can be “as good as random” when done properly.

Propensity scoring can be especially useful in retail when a test may require a high degree of operational change or “hands-on” execution. Random selection may mean that stores on opposite ends of the country are part of the test. That means: (1) two separate regional managers need to be trained, (2) media and other marketing materials need to be adjusted in multiple regions, and (3) inventory and merchandising need to be considered in multiple distribution networks. Propensity score matching could be used to select stores within regions, reducing operational costs of running the test.

At one large retailer, management insisted on launching a new operations compliance program in the worst performing stores. They challenged us with “You recommended the program; it is going to work, right?” Even in this case, we were able to find some similar stores not receiving the program. These stores served as a decent control group and allowed us to measure the benefit of the program.
While we discuss test and control in a brick and mortar setting, these methods are even more commonly used by online retailers (often in the form of A/B testing) due to the low cost of creating test and control groups. Omnichannel retailers can and should employ these methodologies across appropriate channels.

**Test, learn, adapt**

We challenge retailers to take a more thoughtful, scientific approach to testing and measurement. This begins by building a culture of measurement. In the more advanced retailers, this culture manifests in a commitment to “getting things right” up front, a thirst for details, and the patience necessary to methodically evaluate operational tests before moving forward. Once this culture is in place, retailers must focus on crafting experimental design and building a repeatable process to measure accurately. This design is even more important in the world of big data, where retailers collect terabytes of data on stores and customers every day. With a little analysis and planning, the data can be put to work effectively. Finally, rather than measuring success through a single metric, we challenge retailers to pay careful attention to the definition of success.

Though establishing a data-driven culture of measurement is no easy task, retailers—such as the one highlighted below—may have the opportunity to fully capitalize on winning pricing strategies.

After all, as Edward Deming once said, “In God We Trust. All others bring data.”

---

**Establishing Test and Control success at a multi-billion dollar specialty retailer**

In 2014, one specialty retailer was beginning to see signs of collapse—share price had tumbled 50 percent over three years, traffic and item sales were down, all while holding average price per item constant. The question was: Should they make pricing adjustments to be more competitive and relevant to its customer base, or would that cause further erosion?

In the past, the answer to that question would have been to launch nationwide price changes based on limited data and gut instinct honed by decades of retail experience. In a reversal from precedent, the retailer decided to take a more measured approach and pilot a large number of pricing tests across its 2,500 stores, with the ultimate goal of transforming the organization’s approach to strategic pricing.

As we began to design and implement the pricing tests, we focused first on accurately assigning test and control stores. Executional considerations were particularly important at this retailer, and tests were generally executed on a district level basis (i.e., eight to ten stores). Due to this constraint, we followed a simplified propensity matching method, using three key variables (revenue, customer activity, and profitability) to match test and control stores. Matched stores came from “clean” districts, where other tests were not currently happening.

Following the test execution, we placed equal importance on success measurement, creating detailed scorecards containing five to ten key metrics to accurately evaluate results. These scorecards, and the attention to multiple measurement criteria, allowed management to understand the power of the scientific process. The program not only furthered the culture of measurement at the retailer, it also helped narrow and prioritize the list of potential pricing actions to be executed going forward.
Anna Hong
Senior Manager, Monitor Deloitte
Deloitte Consulting LLP
annahong@deloitte.com
Anna Hong is a senior manager with Deloitte Consulting LLP and a member of the firm’s Retail Pricing practice. She has been with Deloitte for over ten years and focuses on strategic and tactical problem-solving around profitable growth opportunities. Her relevant experience spans work in North America, Europe and Asia, and most recently, Anna led Deloitte’s Pricing Center of Excellence, focused on developing methods, tools, and learning courses (she has also been teaching pricing courses at the undergraduate and graduate levels). Anna holds a BBA from the Goizueta School of Business at Emory University and an MBA from the Kellogg School of Business at Northwestern University.

Drew Gaputis
Senior Consultant, Consulting
Deloitte Consulting LLP
dgaputis@deloitte.com
Drew is a senior consultant with Deloitte Consulting LLP in the Strategy and Operations practice. He has experience in consumer products industries with a focus on pricing profitability management and corporate growth strategy. He has primarily served retail clients helping to lead value-based pricing transformations and new market growth strategies. Prior to joining Deloitte, he had more than four years of professional experience working as an equity research analyst within the consumer products, media, and information services sectors.

Kim Frazier
Principal, Monitor Deloitte
Deloitte Consulting LLP
kfrazier@deloitte.com
Kim is Deloitte Consulting LLP’s Retail Pricing Practice leader, bringing with her twenty years of consulting experience in the retail, consumer services, and high tech industries. Kim has led several successful transformational pricing programs centered on profitable growth, omnichannel price strategy, and business process redesign. Kim’s project teams develop strategies to solve our clients’ sales and profitability challenges while leveraging advanced analytics to uncover significant, actionable insights. During the course of her career, Kim’s teams have delivered over half a billion dollars’ worth of benefits to her clients. In addition to her focus in pricing, Kim leads various merchandising, marketing transformation, channel strategy, sales compensation alignment, and omnichannel projects.

Michael Greene
Senior Manager, Consulting
Deloitte Consulting LLP
migreene@deloitte.com
Mike is a senior manager with Deloitte Consulting LLP and a member of the firm’s Advanced Analytics and Modeling practice. For over ten years, he has designed, built, and implemented advanced analytics and predictive modeling solutions to help companies across industries improve business processes, operations, and profitability. Leveraging predictive analytics, optimization techniques and machine learning solutions, Mike has assisted retailers, consumer products manufacturers, technology companies, insurers, and public sector entities drive significant savings and improved profitability. Mike holds an AB in mathematics from the University of Chicago, and an AM in statistics from Harvard University. He has been a member of the American Statistical Association since 2003 and is based with Deloitte in the Boston office.
Works cited