



Case Study: A-maize-ing Retail Marketing-Mix Test *22% jump in retail sales after one test of 10 in-store advertising, display, and promotional elements*

Introduction

Supermarkets are a battleground for packaged goods. Competition for shelf space and the cost of in-store promotions make retail a tough market for even the largest companies. A-maize-ing Foods*, a small producer of corn chips and snacks, had some initial success with their unique products and packaging. Their blue corn chips and sweet potato and other vegetable chips had solid sales and growing acceptance among supermarket chains and convenience stores.

However, A-maize-ing Foods' president, Jonathan Gibson, worried about market pressures as their sales grew. Their foods needed the visual appeal and unique flavor to entice people to buy a bag and keep coming back for more. The company also had to fight for shelf space with penny-pinching retailers. And greater market exposure increased the threat from competitors who had large marketing budgets and skill in knocking-off successful brands.

The company regularly tested new products, prices, and promotions. One winner was their A-maize-ing Blue Corn Chips, a new product in the segment for healthier natural foods. The chips came in a bright-colored bag with a quirky creative image and the chips had a unique spicy flavor that proved to be popular.

The management team believed this product (that they nicknamed ABC Chips) had immense potential, so they decided to test new ways to increase sales and profitability. The team had a long list of marketing-mix changes they wanted to test, but money was tight, so Jon decided to keep the test small. After some heated debate about which one or two changes they should test, the director of database marketing suggested an alternative. He suggested that they use advanced testing techniques to give them the freedom to test more variables at the same speed and cost of testing one or two variables alone. After agreeing on the approach, Jon hired an expert from LucidView to guide them through the project.

** Note: the company name and certain details have been disguised to protect proprietary information, yet the results remain true to the actual case study.*

10 Elements in One Retail Test

Since every retail display, advertisement, and promotion cost money, the team wanted to find the few most profitable changes to implement. After brainstorming 32 marketing-mix changes, the team identified 10 elements they wanted to test.

| <u>Test Element</u> | <u>(-) Control</u> | <u>(+) New Idea</u> |
|--|-----------------------------------|----------------------------------|
| A: Display in produce (by guacamole) | No | Yes |
| B: Rack by beer / seasonal display | No | Yes |
| C: Add product on natural food aisle | No | Yes |
| D: Cross-promote with salsa | No | Yes |
| E: Shelf position | Eye level | 1 shelf down (cheaper) |
| F: Packaging | Quirky image with see-through bag | "All natural" focus in solid bag |
| G: Discount | Low | High |
| H: Advertisement on grocery divider | No | Yes |
| I: Ad in store circular (regional block) | No | Yes |
| J: On-shelf advertisement | No | Yes |
| K: empty | | |

Most of the elements were simply yes/no changes, testing the benefit of having an additional rack or retail promotion. For other elements, they selected two fairly bold changes to test, like a completely new bag or a higher discount. The elements included:

A: Display in produce

Supermarket research has shown that people usually go around a grocery store in a counterclockwise direction, starting on the right side (usually where produce, deli, and fresh food is located) and circling around. With the idea of catching customers *before* they reach the snack aisle, the team tested a separate display in the produce section on top of the stand-alone refrigerated area where guacamole is sold.

B: Rack by beer / seasonal display

At the opposite end of the store, supermarkets usually sell beer and sometimes seasonal items. Since chips and beer have always been a culinary match, the team decided to test a display rack close to the beer. Also, this would give customers one more chance to buy chips before check-out.

C: Add product on natural food aisle

The supermarket chain used for the test had a separate “natural food” section a few aisles away from snack foods. This section included a wide range of foods from organic milk and tofu to dried fruit and novel snacks. Since their ABC Chips were classified as organic, they thought additional shelf space in this section might increase sales to a segment of customers who may avoid the regular snack food aisle.

D: Cross-promote with salsa

The final display location they wanted to test was by salsa, also a few aisles away from snack foods. As a natural match, they tested a small stack-out display (a rack in the aisle with bags hanging from clips).

E: Shelf position

In the main snack aisle—the largest area where ABC Chips are sold—the company paid a premium for the best shelf space (experts say people are more likely to buy from eye-level). The team wondered if a cheaper shelf position, one shelf down, would have any impact on sales.

F: Packaging

ABC Chips came in a brightly-colored bag with a quirky A-maize-ing logo and graphic and a see-through window showing the purple chips inside. The team had never tested the bag, so they wondered if a new look with an “all-natural” theme, less-quirky graphic, and foil-lined windowless bag (like other potato chips) might appeal to a larger segment of the market.

G: Discount

A-maize-ing’s marketing team knew that the 10% discount they often ran increased sales and profitability, but they did not know if 10% was the optimal level. So for this test, they wanted to double the discount to 20% and see if the increase in units sold would make up for the lower margin.

H: Advertisement on grocery divider

One creative new idea they brainstormed was to place an ad on the four sides of the plastic stick-like grocery dividers people use at checkout to separate one person’s groceries from the next. They found that the supermarket would not charge very much for this advertising, so they produced custom ABC Chips grocery dividers for the stores to test.

I: Ad in store circular (regional block)

At the entrance of many supermarkets there is a rack with a stack of newspaper circulars showing the weekly specials. The team wanted to test the value of having an ad for ABC Chips in the store circular—not only in-store, but also as a part of the Sunday newspaper insert. Since the same newspaper circular goes out to everyone in a particular region, this element had to be a regional “block.” For the test, all stores in one region got the circular (I+) and all stores in another region did not (I-). Therefore, stores from two different regions had to be included in the test and, statistically, the effect of the circular could not be separated from the region-to-region difference in sales.

J: On-shelf advertisement

Supermarkets offer a few options for placing ads in the store—with signs, coupon dispensers, and even floor graphics. The team decided to test ads attached to the shelf at the primary location in the snack aisle. These small postcard-size ads are in a plastic border that snaps onto the shelf.

K: empty

This test design allows for up to 11 elements, but the team only wanted to test 10, so they just left one column empty.

The Multivariable Test Design

With these 10 test elements, the statistical consultant used a 12-recipe “reflected” Plackett-Burman design (with 24 total runs). The Plackett-Burman design was used to minimize the number of test recipes (“test cells”). A 32-run fractional-factorial design in most cases would be a better choice, but with a limited number of stores available for testing (as explained below), the 12-recipe reflected design was chosen instead.

The “reflected” design (also called full-foldover) was selected to eliminate the confounding of main effects with 2-way interactions, changing the design from Resolution III to Resolution IV. The term “reflection” is used because twelve additional recipes are run with every “+” and “-” switched, somewhat like holding a mirror up to the original design. For example, recipe #2 is: A+, B+, C-, D+, E+, F+, G-, H-, I-, J+, K-. For the second reflected recipe, #14, all signs are reversed to become: A-, B-, C+, D-, E-, F-, G+, H+, I+, J-, K+. The reflected design doubles the number of required combinations, but eliminates 2-way interactions from the calculation of main effects and helps pinpoint potentially significant 2-way interactions:

1. Main effects – analyzing all 24 test recipes in each column gives a more accurate measure of main effects, independent of 2-way interactions.
2. 2-way interactions – analyzing the first 12 recipes in each column and the second 12 recipes separately, any significant difference in effects is due to one or more 2-way interactions.

Because of the time and cost of producing many combinations, reflected designs are seldom used in direct mail, print advertising, or even Internet applications. But for retail, additional recipes add little, if any, additional cost. Each store needs to be set up and monitored individually, so more stores require more effort, but the number of unique recipes doesn’t really make a difference. The statistical benefits far outweigh the cost of implementation. The only constraint is the number of test units available (i.e., the number of stores that can be used for the test).

Three of the test recipes are shown below.

| Test Element | Recipe #1 (control) | Recipe #2 | Recipe #3 |
|--|------------------------|----------------|----------------|
| A: Display in produce (by guacamole) | No | Yes | Yes |
| B: Rack by beer / seasonal display | No | Yes | No |
| C: Add product on natural food aisle | No | No | Yes |
| D: Cross-promote with salsa | No | Yes | Yes |
| E: Shelf position | Eye level | 1 shelf down | Eye level |
| F: Packaging | Control bag | New, solid bag | New, solid bag |
| G: Discount | Low | Low | High |
| H: Advertisement on grocery divider | No | No | Yes |
| I: Ad in store circular (regional block) | No | No | No |
| J: On-shelf advertisement | No | Yes | No |

Recipe 1 was simply the control. All other recipes had about half the elements set at the control level and half at the new level, but a different half and half for each recipe. Though these three may look like random combinations, all recipes fit within the precise statistical test design. Like the pieces of a puzzle, all recipes fit together to provide accurate data on the main effects and important interactions of all 10 elements.

Statistical Details: Key Metrics and Test Units

The key metrics for this test were sales and gross margin—testing to see if each element increases sales and whether that increase covers the cost of the display or promotion. However, retail tests have some key challenges over other marketing tests. Since relatively few stores are used for the test and each store has a different historical sales level, the key metric is actually the change in sales versus the predicted sales per store. For example, store #30 may sell about 100-150 bags of ABC Chips each week, but store #40 may sell 200-260 bags each week. So sales during the test period must be compared to an average of 125 bags/week for store #30 and 230 bags/week for store #40.

Calculating the baseline sales level for each store can be complicated and potentially a large source of error. If stores vary widely in sales levels, then they should not be grouped together in the same test (since confidence in a 10% changes in sales is much different for a store that sells 10 bags one week and 11 the next, versus a store selling 1000 bags one week and 1100 the next). Also, “special causes” may have a big impact if, for example, a store in Ocean City is tested in June, or a store in Chicago is tested during a January snowstorm.

Advanced statistical techniques can be used to model each store’s sales using months or years of historical data, but more straight-forward calculations can be equally accurate. In this test, sales were averaged for the five weeks prior to the test period and compared to sales during the test in order to calculate the key metric of percent change in sales. With these sales data, gross margin was calculated by subtracting the weekly cost of the display or promotion (in this case study, only sales data are provided).

Unfortunately, in this case, sample size was determined more by the marketing schedule and budget rather than statistical requirements... and this case shows how the details of test execution are often far more challenging than the statistics.

Test Units

The consultant suggested a minimum of 96 stores for the test and, ideally, 96 stores in each of two different supermarket chains. With a resolution IV 32-run fractional-factorial design, this would have given three stores—three replicates—per recipe (x2 chains), so an outlier within each recipe would be easy to identify, plus a measure of similarities and differences between supermarket chains. However, analyzing the cost of the test, company management set the limit at 50 stores within one supermarket chain. The consultant did not want to risk having just one store in some test cells, so he changed the test to a 12-recipes reflected design with just two replicates in each recipe.

The consultant analyzed sales data for all stores in the two regions used for the test. Eliminating special causes—like new stores and stores with strong trends up or down—he took the largest 24 stores in each region, then listed them in order by sales volume. He then grouped one of the largest with one of the smallest stores and went on down the list until all stores were paired up. This way, each recipe had one large and one small store, but combined together, historical sales were about equal for each. Statistically, this gave just one test unit (of two stores), with each week defined as replicate, so week-to-week variation would be used as a measure of experimental error. Analyses were also run separating out week-to-week plus store-to-store variation. With a fixed “week” effect and two stores nested within each recipe, four data points (2 weeks x 2 stores) could be analyzed per recipe. This approach showed a larger measure of experimental error, but did not change results from those shown below.

Schedule

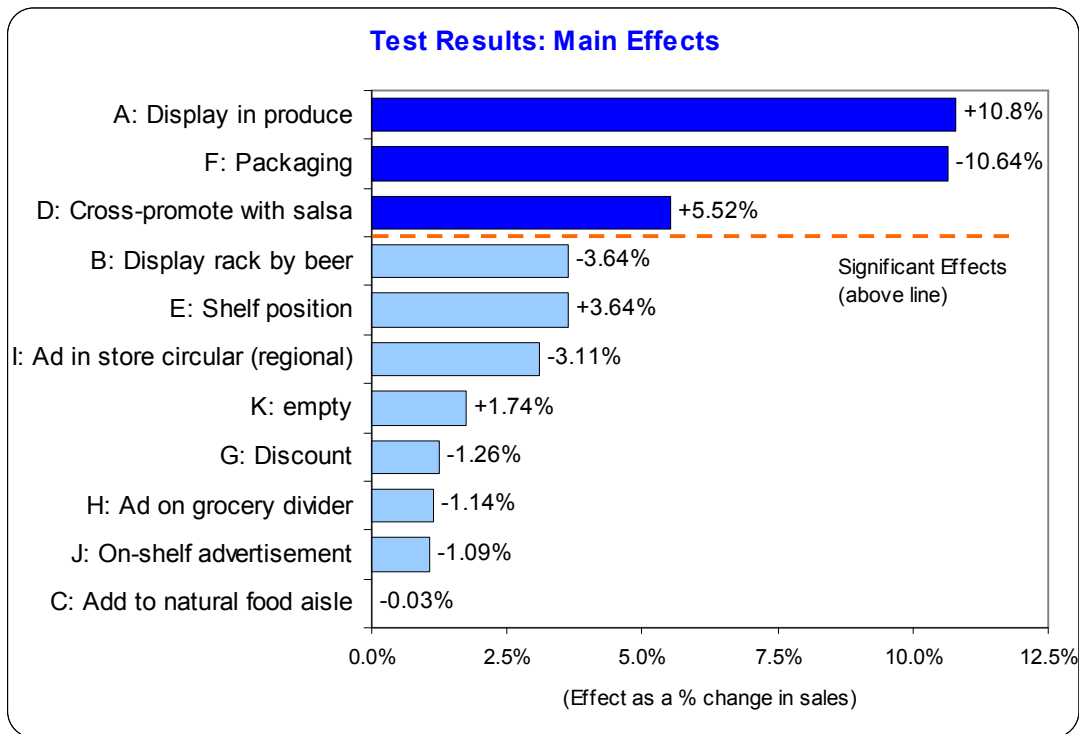
With weekly fluctuations in sales and fewer stores, the consultant suggested two months for the initial test. But costs, deadlines, and delays took a toll. All of these elements were fairly straightforward to execute. Once the displays, promotions, prices, and packages were in place at each store, they were easy to monitor. The consulting team planned to explain the test to all store managers and assistant managers and visit each store twice a week to ensure compliance. But company management was concerned about the cost of testing four displays, three in-store ads, and a higher discount. Though the long-term benefit far outweighed the short-term cost, management wanted to reduce the number of weeks the test ran as much as possible.

Jon Gibson set a deadline of mid-May for all testing to be completed, so results could be implemented at the start of the highly-profitable summer season. As the marketing group worked with vendors and supermarket executives to create each test element and get approval for the test, the schedule fell far behind. Everything was not in place until mid-April.

Everyone was excited to kick things off when the rack vendor—who promised to deliver all racks to the stores over the weekend—called Jon with news that the racks were not ready. Jon met with his management team and the consultant and discussed their options. After all their work, they decided to go ahead with the test, but not until the racks were in place. Finally, the test began at the end of April—leaving just two weeks for the test to be completed.

Test Results

Analyzing results from all 24 recipes (48 stores) over two weeks, the consultant calculated all main effects, summarized in the bar chart below.



With less-than-ideal conditions and just a two-week test, these results gave Jon Gibson and his marketing team all the information they needed.

Three effects were strong:

1. A+: Display in produce (by guacamole)

The display on top of the refrigerated case in the produce section (where guacamole is sold) increased sales 10.8%. This not only identified the one most profitable new location, but also supported their theory that catching the customer early will increase sales. This was a big change, placing ABC Chips away from standard snack food areas.

2. F-: Packaging in the original quirky and colorful bag

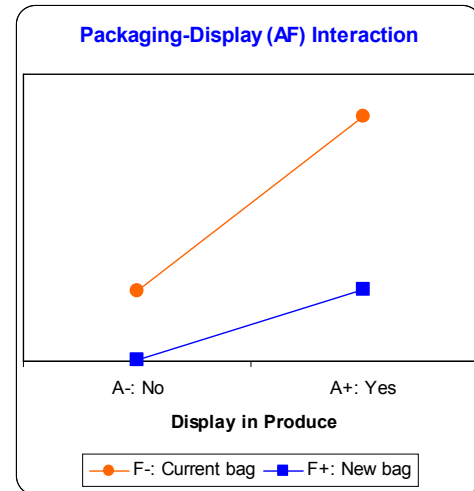
Sales dropped by 10.6% with the new, more common-looking bag with an “all natural” theme. Not only did they prove that their original packaging was a winner, but it also helped them realize that they needed to maintain the unique and “quirky” image in the marketplace.

3. D+: Cross-promote with salsa

A stand-alone display by the salsa (a few aisles away from snack foods) was the next largest effect, increasing sales 5.5%. This was a talk rack in the aisle with bags hanging from clips. It was not something they would use all the time, but perhaps every few weeks or with a special promotion. This display supported the same theory they had about the produce display—that the chips should be located with related foods outside of the highly-competitive snack food aisle. Yet this option was much cheaper than the end-cap displays (large displays at either end of an aisle) that larger competitors favor.

In addition, an important two-way interaction provided deeper insights into these key elements.

The line plot, above, shows the AF interaction. The main effect of A (display rack in the produce section) changes significantly depending on the packaging (element F). As shown in the bar chart, the display in produce is always helpful (going from left to right) and the current “quirky image” clear bag is better than the “all natural” positioning (top line versus bottom line), but the impact of the produce display is much greater with the current packaging (A+F-, upper right point). This interaction shows that (1) packaging is an integral component of product displays and (2) ROI calculations will be more accurate when the interaction is considered along with significant main effects.



Conclusions

Multivariable vs. Champion-Challenger Testing

In this case, what was the advantage of using multivariable testing? Well, if the team had used simple champion-challenger techniques ...

- Testing 10 elements in 48 stores, not one effect would have been significant, since the line of significance would have been 3 times higher (only effects >15% would be significant)
- For equal confidence, the team would have to use 316 retail stores instead of 48
- The team would have never uncovered the AF interaction

21.8% Increase in Sales

Overall, the average sales increase during the test was 8.3%. Adding these three significant effects (calculated as the overall average plus ½ of each effect) gave a sales increase of 21.8% versus the five-week baseline (even more impressive compared to the “control” test cell, #12, which showed a 7% decrease in sales).

The non-significant effects were also very valuable. With such a brief test, it can be risky to assume non-significant effects have no impact on sales, but these results do show where the company can get the largest return on their marketing dollars. Jon and his team decided to avoid paying for the other non-significant displays and advertising, though they might test them again after the summer season. The larger discount did not have a big impact on sales, so they decided to stick with a 10% discount (when offered). Although shelf position was not significant, they were cautious about changing to the less-prominent position without further testing, so they kept the premium, eye-level, shelf space.

By the end of May, Jon Gibson and his team at A-maize-ing Foods had everything set for the summer rush. They increased sales dramatically while keeping costs low and focusing every marketing dollar on maintaining their unique position and grabbing more customers. Jon still was not taking their success for granted, but he relaxed just a little—knowing that with discipline, experience, and continual testing he could hold his own in a tough marketplace.

© LucidView 2006. All rights reserved.