2003 Advertising Research Federation
Annual Convention

Technical Paper for Conference Session
April 10, 2003

Innovations in Retail Testing and Measurement:
Testing a Multitude of Marketing Mix Elements All at Once

by Gordon H. Bell

Gordon H. Bell
LucidView
888-lucidview (888-582-4384 x707)
gbell@lucidview.com

Innovations in Retail Testing and Measurement: Testing a Multitude of Marketing Mix Elements All at Once

by Gordon H. Bell

Abstract

Market testing is essential in quantifying real-world performance. Yet as back-end data collection and analytical systems have improved considerably, front-end test design remains mired in techniques of the 1920s. However, in the last few years, industry leaders have begun to apply advanced market testing techniques that give them the freedom to test dozens of variables all at once as fast as a simple one-variable test. As the science and strategy have been streamlined, these advanced techniques have become more effective, yet remain largely unknown to most marketers, statisticians, and business leaders. This paper explains these advanced market testing techniques, including: (1) the statistics behind scientifically-valid multivariable test designs, (2) a strategy for applying these advanced techniques in market testing, and (3) one example of a retail packaged-goods test using this approach.

Introduction

Market Testing has broad significance in a wide array of fields. Advertisers test their return on radio, television, and print advertising; brand managers test new products, retail displays, price points, and media mix; and direct marketers frequently test new catalog covers, e-mail offers, direct-mail packages, and Internet advertising. In measuring actual purchase behavior, market testing remains integral to the successful launch and continual optimization of new products and marketing programs.

However, this widespread use of market testing masks the reality that the practice has fallen far behind the science. Since the 1920s, test designs and statistical techniques have evolved, but marketers have held firmly to the "champion-challenger" approach (a.k.a. A/B split, test-control, yes/no test), testing one variable—or one new ad, product formula, or mail piece—against the "control," or status quo. Though this approach is completely valid for testing one variable alone, it is just one option out of a wide range of test designs and strategies.

After seeing impressive results using advanced testing techniques, a few industry leaders are beginning to change the perception and practice of market testing. These marketers are using techniques—often called "matrix" or "multivariable" tests—that give them the creative freedom and scientific precision to increase the speed, scope, and clarity of testing: (a) testing three or thirty-three variables with the same sample size, just as fast, as a standard one-variable test; (b) testing a multitude of product, packaging, pricing, and promotional elements all at once; and (c) quantifying the precise impact of each element and interactions between elements. Marketers using these techniques may learn in 3 weeks what might take 8 or 10 months using the standard one-variable-at-a-time approach.

This new strategy for market testing requires unique test designs and careful statistical analysis. Along with the science, these techniques also require greater discipline in the "art" of designing tests and selecting, defining, and executing potent test elements. However, as advanced market testing has been refined and streamlined over the last few years, the scientific complexity has also become more transparent. These techniques require skill and experience, yet have proven robust and efficient for fast, real-world market testing.

The Evolution of Market Testing

When Claude Hopkins wrote <u>Scientific Advertising</u>[1] in 1923, he introduced advertisers to then-current statistical testing techniques. Yet as statisticians—like Sir Ronald Fisher in his famous 1935 book, <u>The Design of Experiments</u>[2]—extended the practice to include "factorial" tests of two or more variables, marketers remained more entrenched in the one-variable-at-a-time approach.

By the mid-1940s, statisticians including R.L. Plackett and J.P. Burman[3] and, more recently, G.E.P. Box, W.G. Hunter and J.S. Hunter[4], advanced the science of test designs to include "fully-saturated orthogonal arrays." Based on complex mathematics and statistical theory, these designs balance the competing objectives of (1) testing many variables at once, (2) minimizing the number of test cells, and (3) running fast tests. However, the statistical complexity of these techniques, the lack of statistical training in business, the unique challenges of testing dynamic markets, and the general lag between theory and practice have led to the continued focus on one-variable tests even as the science has evolved considerably.

Though superficially they may appear incompatible with the "scientific method," these advanced techniques are statistically valid. The important concept in the scientific method is the ability to isolate the effect of the variable you are studying. If a lot of other things are changing at random, or at the same time, then the impact of that one change cannot be isolated. The key to this advanced market testing strategy is that you can change many variables at once in a scientifically-valid test and still isolate the impact of each.

The Simple Science Behind the Complex Techniques

Understanding advanced market testing techniques starts with the acceptance of simple two-variable tests. For example, say an advertiser wants to test a new headline in a magazine advertisement. He can design a basic split-run test:

1. Create two advertisements, each with a unique coupon or 1-800 number (or any other direct-response tactic)
2. Select one or more magazines for the test
3. Randomly run the ad with the "current headline" (i.e. the "control") in half of the copies and the ad with the "new headline" (i.e. test version) in the other half
4. Compare average response for both ads to calculate the effect of the new headline

Now, what if this advertiser wants to also test a new photograph in the ad? He can design a second test like the one above—with a separate analysis and additional sample size required—or he can combine both elements into one test (see Figure 1).

Figure 1

	Headline	Photo
1	current	current
2	new	current
3	current	new
4	new	new

This test design was a major step forward and explains the basic concept behind many advanced test designs. Essentially, it is a test of all combinations of the two test elements. Test cell #1 (also called a test "recipe") is the control. Recipe #2 is the new headline alone, recipe #3 is the new photo alone, and recipe #4 is both the new headline and new photo.

However, combining both test elements into one test design has certain advantages:

1. **Sample size is the same as for a one-variable test**
   For "headline," recipes #1 and #3 are like the control and recipes #2 and #4 are like the test version, so no matter how many variables are in the test, you still have half the customers seeing the test version and half seeing the control—sample size can be the same.
2. **The test is perfectly balanced, so effects can be analyzed as easily as in an A/B split**
   Looking at both recipes with the "current headline" (#1 and #3), one has the "current photo" and one has the "new photo." This means that when you average response across both of these recipes, whatever effect the photo may have just cancels out (whether the difference between current and new is large or small, by averaging, we get the midpoint), so you end up with the effect of the headline independent of the other test element.
3. **The interaction between elements can be calculated since different combinations are tested**
   Each headline (current and new) is tested with both the current and new photo, so you can analyze the interaction between elements to see if the effect of the headline or photo changes depending on how the other element is set. For example, does the new headline work better with the new photo than with the old photo?

The concept of an interaction can be challenging to grasp, yet can lead to valuable insights. For example, recent tests have shown:

- A 15% discount and a free gift each increased sales, but both together were no better than the discount alone
- A dot-whack on the cover of a catalog (and in another test, the sticker on a letter) had a different effect depending on the message on the dot—numbers pulled better than words alone

Full-factorial Test Designs

A test of all combinations of two or more variables is often called a "full-factorial" test design. To find the number of test cells required, you can multiple the number of levels for each test element. For example, the two-level, two-element headline-photograph test requires 2x2=4 test recipes. A test of four elements, each at two levels, requires 16 recipes.

Often, full-factorial tests are run with each element at two levels to minimize the number of recipes, using a "–" to represent the control and a "+" to represent the new level. For example, the headline-photo test can be displayed as in Figure 2.

Figure 2

	Headline	Photo	Response
1	–	–	2.0%
2	+	–	2.1%
3	–	+	2.2%
4	+	+	2.8%

This is the same design as in Figure 1, but with symbols for the levels, and the key metric, response, listed for each recipe. Results can be analyzed using analysis of variance, regression, or frequently with simple mathematics. To calculate the effect of the headline in this example, one can simply take the average of recipes 2 and 4 and subtract the average of recipes 1 and 3.

Effect(headline) = (2.1% + 2.8%)/2 – (2.0% + 2.2%)/2 = 0.35%

Therefore, the new headline increases response by 0.35 percentage points over the old headline. The effect of the new photo is the average of recipes 3 and 4 minus the average of recipes 1 and 2, or 0.45 percentage points.

To see the headline-photo 2-way interaction you can plot all four points, as shown in Figure 3 (you can also calculate the interaction, but we won't get into that here).

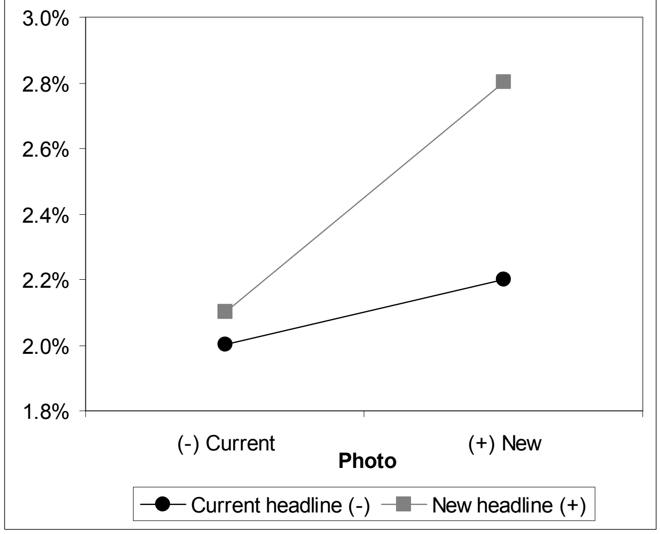


Figure 3: 2-Way Interaction

In this interaction plot:

- The two points on the left show response rate for the two recipes with the current photo
- Both points on the right show results with the new photo
- The top line shows results from both ads with the new headline
- The bottom line shows results from both recipes with the current headline

This plot illustrates the "main effects"—that the new photo (right side) and the new headline (top line) are better—along with the interaction. The interaction shows that the new photo and new headline together are even better than each one on its own. It takes awhile to grasp the concept, but interactions can lead to greater performance and deeper market insights than main effects alone can show. Plus, interactions remain forever hidden in common A/B splits.

Full-factorial test designs like this one (testing all combinations) give you the freedom to:

- Test any number of elements with the same sample size
- Uncover interactions, along with the main effect of each element
- But… each new test element doubles the number of recipes in the test

Certainly this approach becomes unreasonable with more than 4 or 5 test elements. However, it can be very beneficial for small, focused tests like price-offer tests and others where the elements are already known to be important and interactions may be significant.

Multivariable "Matrix" Test Designs

Though more specialized designs exist, the most advanced test designs usually used in marketing combine the strengths of full-factorial tests with the fewer recipes of A/B splits. Essentially, they work like a full-factorial test with many more test elements stuffed into the design.

For example, you can test all combinations of 4 test elements in a 16-recipe full-factorial design. Alternatively, with 16 test cells you can run 15 separate one-variable tests against your control (as the 16th test cell). "Matrix" market tests (also called multivariable test designs) let you test up to 15 elements in one 16-recipe test and analyze the most likely interactions along with all 15 main effects.

Interactions and Confounding

A significant advance in statistical test design was the step from testing all combinations of a few variables to testing few combinations of many variables. To understand how these advanced designs are created, one can start with a full-factorial design (Figure 4).

Figure 4: Full-factorial Test Design

	Headline	Photo	Copy				
Recipe	**A**	**B**	**C**	**AB**	**AC**	**BC**	**ABC**
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

This test includes three elements—A: Headline, B: Photo, and C: Copy—tested via eight versions of an advertisement (created following recipes 1-8). For example, in recipe 1, all elements are at the "control" level and in recipe 8, all elements are set at the "test" level. The four columns beyond column C are all the interactions created by various combinations of the test elements. The interaction columns are calculated simply by multiplying the main element columns together. For example, the AB interaction for recipe #1 is –1 (from column A) times –1 (in column B) to give a +1 (or just "+") in the AB column; the ABC interaction for recipe #4 is (+1)*(+1)*(–1)= –1 in the ABC column. These interaction columns are used to calculate independent interaction effects, just as all main effects are independent of all others.

Now, although all effects can be calculated, the likelihood of significant interactions is lower than the chance of seeing significant main effects. For example, the chance of the headline, photo, and copy all working together to create an impact different than each one on its own is fairly small. Also, higher-order interactions—like the 3-way interaction—are rare and, even if present, are usually small.

Since interactions tend to be small or non-existent, statisticians considered using those interaction columns for additional test elements. For example, placing a new test element, "D: product price" in the AB column, "E: size of ad" in the AC column, "F: page placement" in the BC column, and "G: color in ad" in the ABC column, the 8-recipe design can be used to test seven independent elements instead of only three (see Figure 5).

Figure 5: 7-element Matrix Test Design

	Headline	Photo	Copy	Product Price	Size of Ad	Page Placement	Color in Ad
Recipe	**A**	**B**	**C**	**D** ~~AB~~	**E** ~~AC~~	**F** ~~BC~~	**G** ~~ABC~~
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

In this advanced design, all main effects remain independent of all others, but the interactions do not just disappear—they become mixed in (or "confounded") with the main effects. No analysis can completely separate the main effects from the interactions effects, but generally this is acceptable because:

- Interactions tend to be small, non-existent, or linked to main effects
- Advanced techniques can be employed to eliminate or reduce the mixing of main effects and 2-way interactions
- Test elements can be defined to be independent, so interactions are "designed-out" of multivariable tests

The last point is a very important aspect of the "art" of market testing—defining test elements that are clear, consistent, bold, and independent of all other elements. When interactions are expected to be numerous or large, then a full-factorial test design—or a "fractional-factorial" design mixing some elements and interactions but leaving key interactions independent—can be used instead. In addition, as Hamada and Wu[5] have shown, significant interactions most often include at least one element with a large main effect, so in spite of many potential interactions, the few most-probable interactions can be analyzed.

Ultimately, the statistical complexity should be transparent to the marketing team. Getting the best information from the marketplace is key. Therefore, each test should be designed to see what the brand manager wants to see. These matrix designs are most valuable in offering greater freedom and flexibility in market testing, offering efficient designs for testing one or thirty-one variables, main effects and/or interactions, and pinpointing the optimal combination of marketing elements.

The following case study shows one example of these advanced techniques in a retail test for a consumer packaged-goods company. Please note: proprietary company information—including data and some of the test elements—has been altered or disguised to maintain confidentiality.

Case Study – Advanced Market Testing in Retail

Competition among retailers, manufacturers, and other brands was driving down sales and margins for one consumer packaged-goods producer. While managers had lots of ideas what might help, no one was certain how to boost sales without hurting profitability, so the brand team turned to advanced market testing techniques to help them pinpoint the best ideas.

With guidance from an expert, the brand manager called a meeting with representatives from marketing, merchandising, advertising and promotions, brand strategy, and market analysis where they brainstormed over 75 ideas to increase sales and margins. They then trimmed the list to 19 elements to test in the marketplace.

Using advanced testing techniques, they simultaneously tested all 19 ideas in 20 regions around the country. The test was run for four weeks within 60 stores. In contrast, standard one-variable methods would require 600 stores over four weeks—or those same 60 stores tested over 40 weeks—in order to test all 19 marketing elements with equal statistical confidence (but without any information about interactions). Essentially, they tested 19 different variables in the time it usually takes to test one variable alone.

Figure 6 shows the 19 elements and levels in the test. The "control" level is usually how the element is currently set and the "new idea" is the change from status quo that someone thinks will increase sales or reduce costs.

Figure 6

	Test Elements	(–) Control	(+) New Idea
A	Package design	Standard	Creative
B	Package size ratio (2 sizes)	1:2	1:3
C	Label copy and graphics	Value positioning	Image positioning
D	Sticker on package	No	Yes
E	Price points	Low	High
F	Price spread (between 2 sizes)	Large	Small
G	Discount	Buy 2, get 1 free	$1.00 off
H	Shelf position	Premium	Standard
I	End cap display	Small	Large
J	Stack out aisle display	No	Yes
K	"Shelf talker" in-store ad	No	Yes
L	Floor graphic in-store ad	No	Yes
M	Cross-promotion with related item	No	Yes, shelf rack
N	Print ad - content	Value	Image
O	Print ad - offer	Coupon	Call for free sample
P	Print ad - frequency	High	Low
Q	Direct mail - self mailer	No	Yes
R	Sample in newspaper wrap	No	Yes
S	Market size	Small	Large

An expert worked with the team to create the test design (see Figure 7). In this matrix, each recipe was run in three stores in one region. Each recipe has about half the elements in the control level and about half running the new idea, except for the last, all-minus "control" recipe.

Just as an A/B split requires two cells to test one variable, a matrix market test requires at least one more recipe than the number of test elements (since at least 20 data points are needed to get sufficient degrees of freedom (n-1) to analyze 19 elements). In addition, the number of recipes in a matrix test is always some multiple of four: testing up to 7 elements in 8 recipes, 11 elements in 12 recipes, 15 elements in 16 recipes, and so on. However, since each recipe gives information about every element, sample size is the same whether you are testing three or thirty-three elements.

After analyzing historical sales, selecting stable test units (i.e., stores), and calculating the predicted baseline sales for each, the team worked with their ad agency and the store managers to execute the test.

Figure 7: Matrix Test Design

	Package design	Package size ratio (2 sizes)	Label copy and graphics	Sticker on package	Price points	Price spread (between 2 sizes)	Discount	Shelf position	End cap display	Stack out aisle display	"Shelf talker" in-store ad	Floor graphic in-store ad	Cross-promotion with related item	Print ad - content	Print ad - offer	Print ad - frequency	Direct mail - self mailer	Sample in newspaper wrap	Market size	Change in Sales
Recipe	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	
1	+	+	+	-	-	+	+	-	+	+	-	-	-	-	+	-	+	-	+	10.4%
2	-	-	-	-	+	-	+	-	+	+	+	+	-	-	+	+	-	+	+	12.2%
3	-	+	+	-	+	+	-	-	-	-	+	-	+	-	+	+	+	+	-	-3.7%
4	+	+	-	-	+	+	-	+	+	-	-	-	-	+	-	+	-	+	+	-18.3%
5	-	+	-	+	+	+	+	-	-	+	+	-	+	+	-	-	-	-	+	7.7%
6	+	-	+	+	-	-	-	-	+	-	+	-	+	+	+	+	-	-	+	5.7%
7	-	+	+	-	-	-	-	+	-	+	-	+	+	+	+	-	-	+	+	19.9%
8	-	+	-	+	-	+	+	+	+	-	-	+	+	-	+	+	-	-	-	8.9%
9	+	-	+	-	+	+	+	+	-	-	+	+	-	+	+	-	-	-	-	-8.1%
10	+	+	+	+	-	-	+	+	-	+	+	-	-	-	-	+	-	+	-	8.4%
11	-	-	+	+	-	+	+	-	-	-	-	+	-	+	-	+	+	+	+	20.9%
12	+	+	-	+	+	-	-	-	-	+	-	+	-	+	+	+	+	-	-	-9.6%
13	+	-	+	+	+	+	-	-	+	+	-	+	+	-	-	-	-	+	-	-1.9%
14	+	+	-	-	-	-	+	-	+	-	+	+	+	+	-	-	+	+	-	21.6%
15	-	-	+	-	+	-	+	+	+	+	-	-	+	+	-	+	+	-	-	10.1%
16	+	-	-	+	+	-	+	+	-	-	-	-	+	-	+	-	+	+	+	2.8%
17	-	+	+	+	+	-	-	+	+	-	+	+	-	-	-	-	+	-	+	-9.0%
18	+	-	-	-	-	+	-	+	-	+	+	+	+	-	-	+	+	-	+	13.7%
19	-	-	-	+	-	+	-	+	+	+	+	-	-	+	+	-	+	+	-	1.4%
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.5%

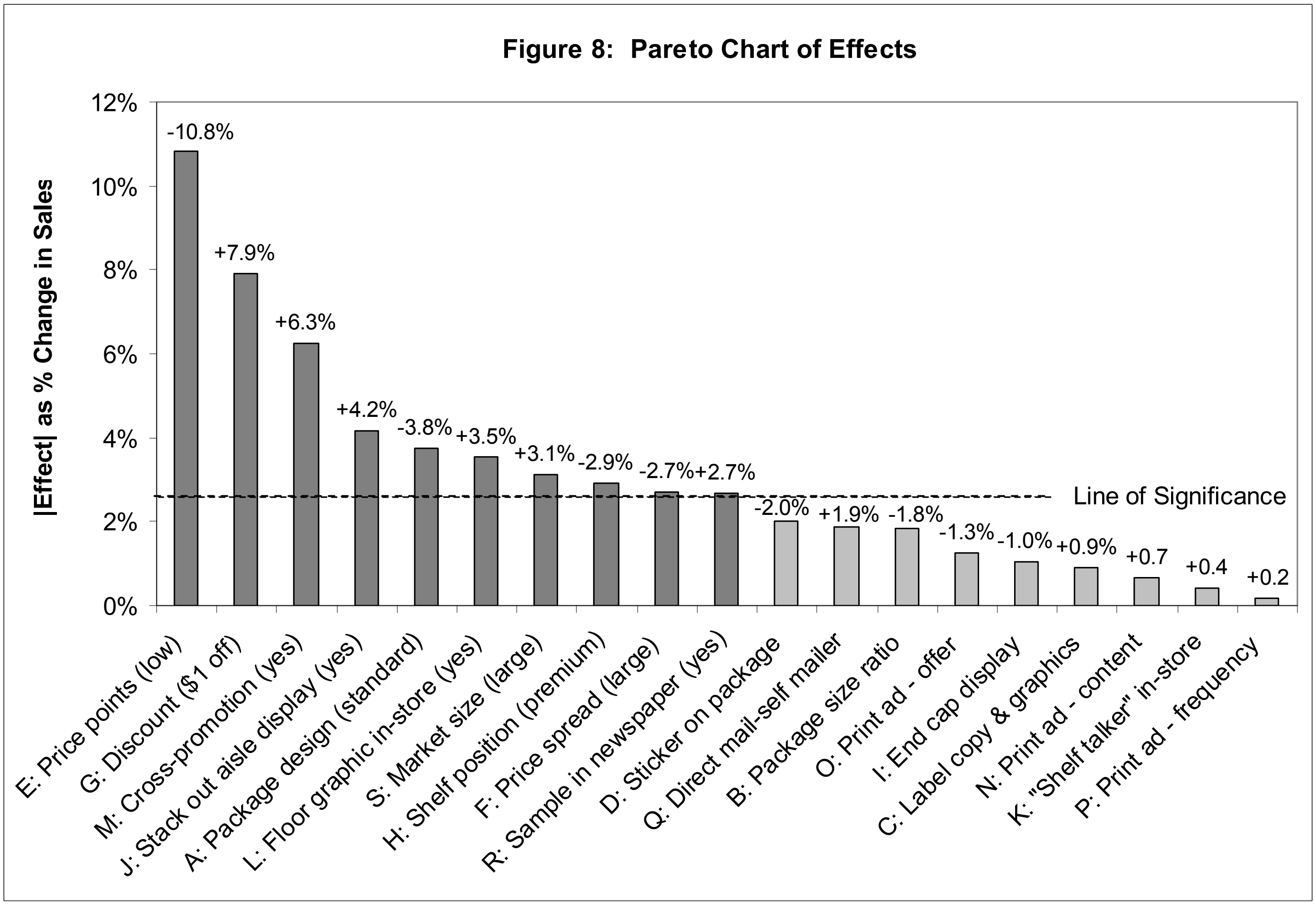
The analysis and interpretation of results is based on advanced statistical concepts. However, a quick analysis of main effects can be conducted with simple addition and division:

Each column is this 20-recipe design has 10 pluses and 10 minuses. For example, looking at column A, recipes #1, 4, 6, 9, 10, 12, 13, 14, 16, 18 have a plus sign. Of these ten recipes, five have a plus in column B (#1, 4, 10, 12, 14) and five have a minus; in column C, a different group of five have a plus and the other five have a minus; and so on for all other elements B-S. What this means is that when you look at the ten recipes with A+, every other element is half "+" and half "–." Therefore, in averaging these ten A+ recipes everything else essentially cancels out, giving you the average sales change due to A+ independent of all other changes going on in the test.

This perfect balance in the matrix (or "orthogonality") means that the main effect of each element is the average of the "+" recipes minus the average of the "–" recipes. Therefore, marketing managers can grasp the test results without relying on a statistician to explain the numbers. However, in-depth statistical analyses are still required to assess things like statistical significance, data reliability, error, and interactions.

This is one significant advantage of a well-designed market test: the up-front effort and discipline in test design, planning, and execution leads to simple, clear results that a marketing team can easily understand and rapidly implement with confidence. As the test process moves forward, the opportunity to overcome a weak design diminishes rapidly as the marketing mix is executed. In other words, market data is a fait accompli; once sales start—once the products, pricing, and media mix are launched—there is no way to go back and correct for a weak test design or otherwise improve the quality of data. No amount of data-mining, modeling, or simulation will uncover answers which the data are not capable of showing. For example, if television, radio, and print advertising start and stop at the same time, there may be no way to assess whether an increase in sales is due to one or all of the media placements.

All main effects can be summarized clearly in a Pareto chart (see Figure 8). The Pareto chart shows all effects from largest to smallest, quantifying the main effect and showing the optimal level for each. The "Line of Significance" is a measure of experimental error: all effects above the line are statistically significant, while all effects below the line can be explained fully by the natural variation in the marketplace.



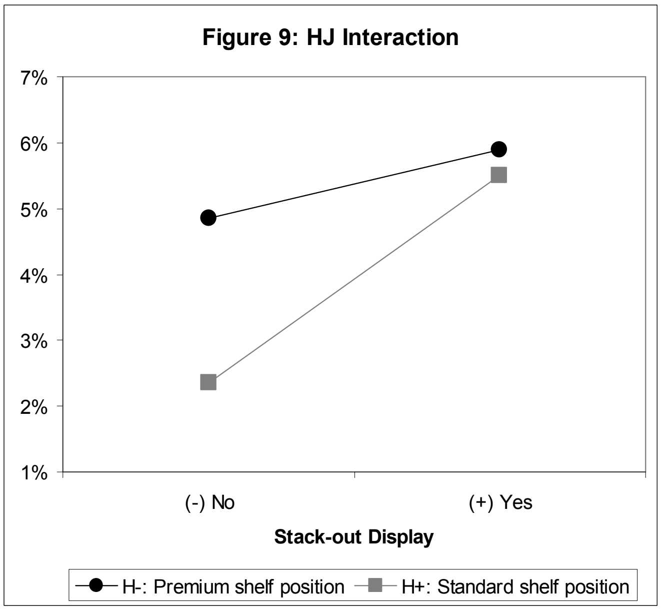
The significant effects range from a 10.8% change in sales between low and high price points (element E) to a 2.7% increase in sales with a product sample inserted in the daily newspaper (element R). The height of the bar (labeled as a percent change in sales) shows the magnitude of the effect and the sign (+ or -) shows which level leads to an increase in sales; the optimal level is in parentheses after each element.

For example, the largest effect, "E: Price points (low)," shows that when prices were changed from "low" to "high," sales dropped by 10.8%. Therefore, this effect is optimal at the minus level (low prices), avoiding a 10.8% loss of sales. The second largest effect, "G: Discount ($1 off)," shows that the new (+) level increases sales over the current (-) level—offering "$1.00 off" leads to a 7.9% increase in sales over the "Buy 2, get 1 free" offer. Overall, six new ideas increased sales, but four new ideas hurt sales (the control level is better).

Non-significant elements can be valuable as well. In this test, a few elements that affected marketing costs had no significant impact: D: Sticker on package, Q: Direct mail campaign, I: Large end-cap display, K: "Shelf Talker" in-store ads, and P: High print advertising frequency. Therefore, eliminating these costly elements should increase profitability. However, although these elements were not statistically significant, they may still impact sales—just not enough to rise above the market noise in this test—so further analysis may be warranted before choosing the low-cost alternatives.

In addition, since the recipes include various combinations of elements, interactions among elements can be analyzed. Though main effects tend to be larger and much more prevalent than interactions, interactions often lead to a greater depth of market insight and sometimes affect the decision on what is implemented in the marketplace.

One significant interaction in this test involved the shelf position (element F) and stack-out aisle display (element J), shown in Figure 9.



In this interaction plot, the top line shows the effect of premium shelf position (H-) with the stack-out display (J+, the circle on the right) and with no stack-out display (J-, the circle on the left). The bottom line shows the effect of the standard shelf position (H+) with the stack-out (right square) and without the stack-out display (left square). This interaction plot shows that—although H- and J+ are always better—the difference between the premium and standard shelf position is minimal if the stack-out display is used (i.e., the points on the right are very close). Alternatively, the stack-out display may be less cost-effective with the premium shelf position (i.e., the top line has minimal slope). Therefore, even though both main effects are significant, both the stack-out display and premium shelf position may be unnecessary. Depending on costs, one element should probably be chosen over the other.

In summary, this retail packaged-goods test had the advantages of:

1. Sample size and speed—testing 19 elements with the same sample size as a one-variable test means testing was completed ten times faster
2. Statistical confidence in main effects and interactions—data on combinations of elements offer clear, in-depth insights into all effects

3. Comparative analysis—by testing all elements under the same conditions at the same time, results let you quantify and compare all effects
4. More variables and more information—matrix tests offer the freedom to "cast a wide net" for specific answers, since additional elements add little or no time and cost
5. Proven answers—solid test designs show clear, cause-and-effect results rather than uncertain correlations from after-the-fact analyses

The Test Cycle and Refining Test

These advanced market-testing techniques offer marketers fast, clear insights into a multitude of marketing variables and also change the strategy of market testing. Instead of planning a random collection of discrete, often unrelated, one-variable tests, marketers can plan continuous "test cycles" including:

1. One large, fast screening test of many variables (like the 20-recipe matrix above) to rapidly pinpoint the most important marketing variables—screening the few important elements from the trivial many
2. One or more smaller refining tests of a few elements (already proven to be significant) to refine levels and optimize important variables
3. Market roll-out of all significant elements, often with on-going refining and confirmation of effects
4. New test cycles as the product, programs, and the marketplace change

Refining tests follow the same basic principles as screening tests, but with more combinations of fewer elements. Usually, these tests use a full- or fractional-factorial design. The advantage is the depth and accuracy of insights you gain into a few important elements. Therefore, refining is most valuable when you have already screened out non-significant variables and want to optimize the key "response drivers:"

(a) Test significant elements at new levels
(b) Analyze interactions more closely
(c) Add new, related test elements
(d) Confirm and quantify effects with greater accuracy

For the retail packaged-goods test, the brand team selected three elements to include in a full-factorial refining test design. This refining test (Figure 10) was run for four weeks in 48 stores (six stores per recipe) with the following elements:

- Price points (testing the low price from screening ± 10%)
- Cross-promotion (testing "none" versus a bolder display)
- Stack-out aisle display (testing the same levels as in screening)

Refining test results are analyzed using the same techniques as for screening. In this case, the three elements were significant—as in the screening test—but with different effects, since levels of the price point and cross-promotion elements were changed from screening. Also, two significant 2-way interactions showed that the higher price point was less profitable (i.e. the effect of the price change was smaller) with the new cross-promotion and/or the stack-out aisle display. These five effects—the three main effects and two interactions—helped the brand team calculate the optimal marketing mix to maximize profitability.

Figure 10: Refining Matrix Test Design

Recipe	A (Price Points)	B (Cross-promotion)	C (Stack-out)	AB	AC	BC	ABC
1	–	–	–	+	+	+	–
2	+	–	–	–	–	+	+
3	–	+	–	–	+	–	+
4	+	+	–	+	–	–	–
5	–	–	+	+	–	–	+
6	+	–	+	–	+	–	–
7	–	+	+	–	–	+	–
8	+	+	+	+	+	+	+

Summary and Conclusions

Market testing techniques have evolved considerably beyond common one-variable-at-a-time tests. Advanced "matrix" test strategies offer marketers and advertisers greater freedom for flexible designs and rapid tests of a multitude of marketing-mix elements, for results that accurately pinpoint and quantify the effect of each. In addition, the analysis of interactions among variables can lead to greater market insights and profitability than one-variable tests can possibly offer.

The potency of market data solidifies as sales occur; once a marketing program is launched, no amount of analytical prowess can uncover insights that the data is not designed to expose. Therefore, successful market testing demands careful up-front planning and test design along with an in-depth analysis of results.

Though the science behind advanced testing techniques is complex, strategies for designing, executing, and analyzing matrix tests have been streamlined to give managers quick, clear, real-world answers. With skilled assistance, the statistical complexity can remain transparent to the marketing team and the application of these tests can speed market insights with minimal additional effort. Marketers have already confirmed the need for market testing… now it's time to do it more efficiently and effectively using the art and science of advanced "matrix" market tests.

About the Author

Gordon H. Bell is the president of LucidView, a consulting firm specializing in marketing science and strategy. He can be reached at 888-lucidview (888-582-4384 x707) or gbell@lucidview.com.

References

[1]Hopkins, Claude, Scientific Advertising, 1923; rpt. Lincolnwood, Illinois: NTC Business Books, 1996.
[2]Fisher, R.A, The Design of Experiments, Edinburgh: Oliver and Boyd, 1935.
[3]Plackett, R.L. and J. P. Burman, "The design of optimum multifactorial experiments," *Biometrika*, 33, 1946, pp. 305-325.
[4]Box, G.E.P., W.G. Hunter and J.S. Hunter, Statistics for Experimenters, New York: John Wiley and Sons, 1978.
[5]Hamada, M. and C.F.J. Wu, "Analysis of Designed Experiments with Complex Aliasing," *Journal of Quality Technology*, 24, July 1992, pp. 130-137.