Original Article

Discrimination tests: Evaluating context effects and respondent reliability using the switchback experimental design

Received (in revised form): 6th April 2009

Nancy K. Keith

is Professor of Quantitative Business Analysis in the Department of Marketing at Missouri State University and an MSU University Fellow, and did her PhD from Purdue University. Her research, primarily involving statistical modelling for Business and the Biosciences, has appeared in numerous journals such as *Decision Sciences*.

Charles E. Pettijohn

is a professor of Marketing at Missouri State University and did his DBA from Louisiana Tech University. He has served in the Marketing Management Association in numerous capacities, and is currently co-editor of *Marketing Management Journal*. His teaching and research interests lie in the areas of Sales and Sales Management.

Megan E. Keith

graduated cum laude with a BS in Marketing from Missouri State University. Currently, she is a graduate student finishing her MBA with an emphasis on Marketing at Missouri State University. Her research interests lie in the areas of Consumer Behaviour, Ethics and International Marketing.

ABSTRACT As companies attempt the risky process of introducing new products and modifying existing ones, many engage in field experiments in which untrained volunteer respondents are used to assess product characteristics and preference. Because discrimination test designs currently employed by marketing researchers have inherent limitations and can independently lead to different results, examinations of other potentially useful design models is vitally important. This paper introduces the switchback experimental design to the marketing literature as an alternative research design for testing new products. The switchback design is recommended for product-testing situations where context effects and respondent reliability issues are anticipated. The design allows for measurement of the effect of product-testing order on a respondent's cognitive and affective judgments, and provides a direct measure of respondent judgment variability. This paper also presents the results of an empirical evaluation of the switchback design for a discrimination taste test. The switchback design promises more accurate feedback regarding both consumer product preference and discrimination, which should lead to better decisions on both new product development and product modification.

Journal of Targeting, Measurement and Analysis for Marketing (2009) 17, 115–125. doi:10.1057/jt.2009.7;
published online 15 June 2009

Keywords: discrimination; reliability; context; switch back; design; marketing

Correspondence: Nancy K. Keith
Department of Marketing, College of Business Administration, Missouri
State University, Springfield, MO 65897, USA

INTRODUCTION

Consumer acceptance of new or reformulated products and/or services is of vital importance for the growth of healthy firms. As organisations seek to expand and/or modify their product offerings,

one aspect that requires significant assessment is the customer's response to the product. The customer's cognitive and affective judgments regarding new and modified products are partially based on comparisons with existing products. Many of the currently employed research designs to test new products require consumers to evaluate the taste, smell, texture, sound or look of selected attributes. These tests may be employed for numerous purposes, including identifying salient product characteristics for product development and evaluating product characteristics for market testing.¹ In addition, product-testing also facilitates product substitutions and modifications, which may be used to reduce product costs or expand product markets.² Thus, accurate assessments are critical for marketers as they attempt to implement new product development and modification strategies.

Unfortunately, the selection of a specific test is often based on historical evolution, habit or the researcher and manager's familiarity and comfort with a particular procedure.³ Because each of the testing designs currently employed by marketing researchers has inherent limitations and can independently lead to different results, examinations of other potentially useful design models is vitally important. However, it has been stated that 'few studies have systematically assessed the strategies, metrics and methods employed in marketing research'.⁴ This paper introduces the switchback experimental design to the marketing literature as an alternative research design for testing new products and services. It also empirically evaluates the use of the switchback design for product taste-testing where context effects and respondent reliability issues are anticipated.

RESEARCH BACKGROUND

Triangle and preference tests

Marketing research often employs comparative taste tests where one product is repeated in the testing sequence. One such approach is a triangle test, where a respondent is presented with three products, two of which are identical, and is asked

to select the one that is different.^{1,5-9} If the respondent correctly identifies the odd product, the respondent is considered to have discriminated between the products.

Another approach is preference ranking, where the respondent is asked to rank three products, two of which are identical, from most to least preferred.^{1,10,11} The respondent is said to have discriminated if the odd product is ranked either first or last. However, it has been argued that perceptual discrimination and preference are not identical measures, and that perceptual discrimination is not a precondition for one to have preferential discrimination.¹² Thus, issues of concern with triangle and preference tests include (1) reliability; (2) overestimation of preference results; (3) context effects; (4) impracticality of large-scale testing; and (5) use of untrained volunteer respondents.

Reliability

If a respondent repeats the same taste test, the results may or may not be identical. To eliminate random responses, multiple triangle tests, increasing the number of samples or altering the sample ratios have been used to assess respondent discrimination ability or preference.¹³⁻¹⁵ Fatigue may occur with multiple triangle tests as respondents engage in three tastings.² Thus, when the number of trial repetitions and/or the number of samples within a trial increases, the increase in psychophysical difficulty of the task must be weighed against the increase in reliability.¹⁴

Overestimation of preference results

Comparative tests have the tendency to overestimate the preference results because of potential guessing by test respondents. Non-discriminators, those individuals who are unable to determine which of two products is the odd sample, are easily identified. However, adjustments must be made to the number of respondents who correctly identified the odd sample. A certain percentage of those respondents may have chosen the correct product sample by guessing. As a result, adjustments are typically

made by subtracting the estimated number of correct guesses from the total who identified the product correctly.

Context effects

The evaluation of a product may be affected by the context within which the judgment is rendered. Research results may be affected by the order of product presentation, fatigue or aftertaste. According to Day,¹⁶ if the products tested are clearly different, order effects may be minor; however, when the products are very similar, order may significantly affect results. In fact, research pertaining to soft drink taste-testing indicated a clear bias in favour of the first soft drink tasted.¹⁷ Additionally, depending on the design of the experiment and the nature of the products tested, fatigue and aftertaste effects may invalidate results.^{5,14,18} Researchers attempt to minimise the effects of order, fatigue and aftertaste by rotating product order across respondents, and requiring respondents to expectorate the products and/or rinse the mouth with water among products.

Impracticality of large-scale comparative testing

When a taste test involves large-scale comparative testing of multiple products on multiple product dimensions, traditional triangle tests may not be practical. When conducting experiments with a large number of respondents, a variety of products and examination of numerous product attributes, the logistics become overwhelming for both the researcher and the respondents. For example, respondents in a supermarket or convention situation may not be willing to participate in a study that requires repeated testing over an extended time period. Further, large-scale tests may require multiple test days or multiple test locations to complete. Therefore, monadic tests are employed.

In monadic tests, all respondents test all products in a randomised order. However, each respondent judges each product only once. Therefore, respondent reliability cannot be measured. Additionally, context effects such as order of product presentation, fatigue or aftertaste

may distort statistical results and lead to erroneous conclusions about product differences.¹⁷ For example, although with monadic tests one-way analysis of variance can be used to compare product rating means on each product dimension, skilful pre-experiment planning allows the use of the Latin square design. The Latin square design specifies efficient randomisation schemes for monadic tests of eight or fewer products. However, in the Latin square design, the interaction between respondents and order of product presentation is assumed to be zero. If an interaction is present then part of that effect is contained within the product effect, and the product effect may appear large when in fact it is not. For example, if an interaction between respondents and order of product presentation exists because of a context effect such as fatigue, then the Latin square results are biased.^{19,20} With context effects present, the results from one-way analysis of variance may be similarly affected.

Untrained volunteer respondents

Additional confounding effects may be present when untrained volunteers are used in a taste test involving multiple products and product dimensions. However, as described by O'Mahony,²¹ the perception of and ability to detect product qualities may vary widely among respondents. However, there are situations where preliminary information about taste test respondents is simply not available. For example, at a convention or trade show, a food vendor might offer samples of new or existing products to volunteers in discrimination tests. In such situations, the researcher has no information about the respondents other than an assumption that they do not have trained palates and are likely to have different taste discriminating abilities. Such assumptions and situations do not allow the researcher the luxury to employ any blocking effects into the research design, and the use of triangle or preference techniques would be questionable. Correspondingly, it has been suggested that triangle tests should be discontinued in the measurement of preference.¹²

The switchback experimental design

The switchback experimental design for more than two treatments, as described by Lucas,²² is a triangle test that permits the evaluation of multiple products over multiple dimensions when a context effect such as fatigue is present. Within the switchback design, a respondent tests three products in an A-B-A sequence where the first and last products sampled are identical. In contrast to traditional triangle or preference tests, respondents are unaware that two of the products in the triad are identical. The respondents are also unaware that the middle product is the odd one. Therefore, the respondents have no prior information that differences exist between the products; the respondents are not predisposed to guess the correct response.

The switchback design is superior to triangle and preference tests in five ways: (1) The A-B-A sequencing enables each respondent to serve as his/her own control. It permits direct comparison of respondent judgments of the identical product and provides a direct measure of respondent response reliability. With the switchback design, respondents are not aware that two of the product samples are identical. Therefore, the switchback design eliminates the concern regarding overestimation of preference results because of respondent guessing. (2) The switchback design analysis uses the difference in the sum of the first and third responses (judgments of the identical product) and two times the second response (judgment of the odd product). By using differences, if respondents are affected to varying degrees by a context effect such as fatigue, the design controls for this source of variation. (3) The switchback technique allows researchers to collect data from large groups of untrained tasters, such as in convention situations, where triangle or preference tests would be unwieldy. (4) The switchback design requires that respondents be randomly assigned to treatments without regard to respondent similarities or differences. (5) If testing involves multiple days or locations, the switchback design accommodates block effects.

RESEARCH METHODS

To illustrate the application of the switchback methodology, a discrimination taste test was designed to include conditions unsuitable for traditional triangle tests or monadic tests such as one-way analysis of variance or the Latin square. The taste test involved several features: (1) comparative testing of multiple products on multiple dimensions; (2) products that were very similar; (3) context effects because of order of product presentation, fatigue and aftertaste; and (4) untrained volunteer respondents.

Products

Four sweeteners were used to empirically evaluate the switchback design – aspartame, saccharin, sucrose and fructose. All products were readily available brand name sweeteners. The sweeteners were dissolved in distilled water at equivalent dilutions according to manufacturer's directions: 28 g of sucrose, 18 g of fructose, 3 g of saccharin and 3 g of aspartame, respectively, per 24 ounces of distilled water. Theoretically, at these dilutions there was no difference in sweetness among the products. The sweetener solutions were offered in unmarked, opaque plastic cups. In order to maximise context effects, no water for rinsing between sweeteners was provided. Saccharin, which has a strong aftertaste, was included as one of the test products. The respondents were not told the types or the number of different sweeteners in the study, nor were they told that the first and last sweeteners in their respective three product sequence were identical.

Product dimensions

Within the switchback protocol, the respondents rated each product on five attributes – sweetness, aftertaste, artificial taste, high calorie taste and whether they liked the sweetener. The attributes of sweetness, aftertaste and whether they liked the sweetener are frequently included in taste tests. The attributes of artificial taste and high calorie taste were of interest in this study because of the common perceptions that artificial sweeteners taste artificial and that sucrose is much higher in calories than artificial sweeteners.

Respondents

Before participating in the study, respondents signed a consent form that advised respondents with diabetes or phenylketonuria to withdraw from the study. The respondents were not screened for any other conditions that might alter taste acuity. The respondents filled out a questionnaire containing several biographical questions including sex, age, smoker/non-smoker, whether or not they selected diet soft drinks and whether they considered table sugar (sucrose) to be higher in calories or to taste better than artificial sweeteners. The biographical questions were asked to assess respondent perceptions and usage of the test products, and to determine whether specific respondent types (for example, smokers) were equally distributed across switchback triads.

Analysis of the biographical data revealed that 60 per cent of the respondents were male and 83 per cent were non-smokers. Respondents ranged in age from 19 to 60 years; however, 62 per cent of the respondents were within 26–30 years of age. As was anticipated, a large percentage of the respondents (78 per cent) felt that table sugar (sucrose) was much higher in calories than artificial sweeteners. An even larger percentage of the respondents (83 per cent) felt that table sugar (sucrose) tasted better than artificial sweeteners. Additionally, 52 per cent of the respondents indicated that they never selected diet soft drinks. Based on biographical characteristics, respondents appeared to be equally distributed across switchback triads.

Instrument and pretest

The respondents were provided with written instructions describing the taste test procedure. The instructions indicated that the sweeteners were to be evaluated in the order presented, and that the respondents were not to go back and retaste any sweetener tested previously in the sequence. Before administration of the study, the instructions and rating instrument were pre-tested on a group of volunteers. The volunteers reported no difficulties with the instructions, rating instrument or testing procedures.

The respondents rated each product on the five attributes – sweetness, aftertaste, artificial taste, high calorie taste and whether they liked the sweetener – on seven-point bipolar scales. A seven-point bipolar scale was chosen to ensure that traditional analysis of variance under the switchback protocol could be performed. Means were interpreted as if the scale was continuous. For sweetness, artificial taste and high calorie taste, the level of the attribute decreased as the scale increased from one to seven. For aftertaste and liked the sweetener, the level of the attribute increased as the scale increased from one to seven.

Experimental design

As illustrated in Figure 1, the switchback experimental design for four treatments used a complete design with three blocks. The letters A, B, C and D represented the four products tested. For example, respondent 1 tested an A–B–A product sequence, whereas respondent 2 tested a B–C–B product sequence. Twelve respondents

Block 1					Block 2					Block 3				
Subject					Subject					Subject				
Time	1	2	3	4	Time	5	6	7	8	Time	9	10	11	12
Period					Period					Period				
1	A	B	C	D	1	A	B	C	D	1	A	B	C	D
2	B	C	D	A	2	C	D	A	B	2	D	A	B	C
3	A	B	C	D	3	A	B	C	D	3	A	B	C	D

Figure 1: A switchback design for four treatments.
Note: A, B, C and D represent the four products.

Source	Degrees of freedom	Sum of squares
Treatment	$p-1$	$(1/(6np))\sum_k Q_k^2$
Error	$(np^2-(n+2)p)/2$	$(SSTO)-(SSTR)$
Corrected total	$((np(p-1))/2)-1$	$((1/6)\sum_i \sum_j D_{ij}^2)-(M^2/(3np(p-1)))$

where:

p =number of treatments,

r =number of subjects per treatment sequence (replications),

$n=2r$,

$D=Y_1-2Y_2+Y_3$ where Y_1, Y_2, Y_3 are the responses from time period 1,2,and 3, respectively,

D_{ij} =the D value for the j th subject on the i th treatment sequence,

$M=\sum_i \sum_j D_{ij}$,

$Q=(\sum \text{ of the } D\text{'s for subjects receiving the treatment in the 1st and 3rd periods}) - (\sum \text{ of the } D\text{'s for subjects receiving the treatment in the second period}),$

Q_k =the Q value for the k th treatment,

Y_k = k th treatment mean= $Y+(Q_k/2np)$ where Y is the grand mean of the original data (not the D 's), and standard error of the treatment difference= $3s^2/np$ where s^2 =mean square error, and mean squares and F are computed in the usual fashion,

$SSTO$ =corrected total sum of squares,

$SSTR$ =treatment sum of squares.

Figure 2: The generalised analysis of variance for the switchback design.

were needed for one replication of the design. Sixty respondents participated in the experiment, representing five complete replications of the switchback design.

The generalised analysis of variance for the switchback design is presented in Figure 2. As all of the testing was completed in 1 day and at one location, no blocks were required. The study contained four treatments and five replications; therefore, $p=4$, $r=5$, $n=10$, $i=1, \dots, 12$, $j=1, \dots, 5$, treatment degrees of freedom = 3, error degrees of freedom = 56 and total degrees of freedom = 59.

Analysis

Data were analysed for respondent biographical profile, differences among the sweeteners on the

five attributes, among respondents variability, respondent rating consistency and context effects. Differences among the sweeteners on the five attributes were measured by analysis of variance under the switchback protocol using the algorithm of Keith and Williams.²³ Among respondents variability was measured by examining the distribution of judgments along the seven-point bipolar scale for each sweetener on each of the five attributes. Each sweetener had a total of 45 judgments on each attribute with 15 of these judgments representing respondent duplicates.

Respondent rating consistency was measured by comparing the differences in respondents' first and second judgments of the identical product. Both the magnitude and direction

of the differences were examined. If the respondent's first and second judgments of the product were identical, the rating difference was 0. If the respondent's second rating of the product was one point higher than the respondent's first rating of the product, a rating difference of -1 was assigned. Similarly, if the respondent's second rating of the product was one point lower than the respondent's first rating of the product, a rating difference of $+1$ was assigned and so on. The larger the magnitude of the difference, the more inconsistent the respondent's first and second judgments of the product. The direction of the inconsistency is indicated by the sign, that is, $+$ or $-$. Additionally, for the significant product attributes, context effects were analysed by comparing the mean difference ratings for the 12 switchback triads via analysis of variance. If no significant differences among the triads existed, no context effects were present.

RESULTS AND DISCUSSION

Dimensions of the products

As illustrated in Table 1, the four sweeteners were judged significantly different on the attributes of sweetness and aftertaste. On the basis of sweetness, saccharin and aspartame were judged the most sweet-tasting, but not significantly different from one another. Sucrose was the next most sweet-tasting, with fructose judged as the least sweet-tasting. On the basis of aftertaste, saccharin was judged as having the strongest aftertaste. Sucrose and aspartame were

judged as having the next strongest aftertaste, but were judged not significantly different from one another. Fructose was judged as having the least aftertaste.

As all of the sweeteners were dissolved in distilled water at equivalent dilutions according to manufacturers directions, theoretically no difference in sweetness among the products should have been apparent. However, in several pretest evaluations of the sweeteners by the testing personnel, the same sweetness differences as those indicated by the test respondents were found. Saccharin and aspartame were found to taste sweeter than sucrose, which was sweeter than fructose. However, saccharin was known to have an aftertaste. In the pretest evaluations, the testing personnel noted the same aftertaste differences in the products, as did the respondents. As for the other product attributes, artificial taste and high calorie taste, the true product differences under the conditions of this study were unknown.

Differences among products on the basis of sweetness and aftertaste may have emerged because these attributes may have had a more clear interpretation to the respondent than the attributes of artificial taste or high calorie taste. As noted by the pretest personnel, none of the products was particularly tasty when dissolved in distilled water. This fact may have accounted for the finding of no differences on the attribute of *liked the sweetener*.

Among respondent variability

The variability in respondent judgments is illustrated in Table 2. For almost all of the products and attributes, respondent judgments spanned the entire length of the seven-point rating scale. A high degree of variability among respondents' response levels was indicated. This may have been due in part to differences in respondent interpretation of the attributes or rating scale or because of differences in respondent taste sensitivity.

Respondent reliability

The magnitude and direction of the judgment differences on the seven-point bipolar scales

Table 1: Mean sweetness and aftertaste rating of the four sweeteners

Sweetener	Means	
	Sweetness	Aftertaste
Saccharin	2.90 ^a	4.23 ^a
Aspartame	3.61 ^a	3.39 ^b
Sucrose	4.70 ^b	3.39 ^b
Fructose	5.59 ^c	2.56 ^c

Note: ^{a,b,c}Means within a column with unlike letters were significantly different at $P < 0.05$; $n = 30$ for each mean. Sweetness increased as the value of the mean decreased; aftertaste increased as the value of the mean increased.

Table 2: Frequency with which a rating (1–7) was given for the four sweeteners on the five product attributes

Sweetener	Product attribute rating scale						
	1	2	3	4	5	6	7
<i>Sweetness</i>							
Sucrose	3	4	4	11	8	11	4
Fructose	0	0	5	12	11	10	7
Saccharin	6	8	12	9	4	5	1
Aspartame	4	11	6	5	6	8	5
<i>Aftertaste</i>							
Sucrose	8	9	6	9	6	5	2
Fructose	14	7	10	4	7	3	0
Saccharin	5	10	7	6	6	7	4
Aspartame	11	8	3	5	8	6	4
<i>Artificial taste</i>							
Sucrose	4	5	8	8	9	7	4
Fructose	3	8	10	12	4	4	4
Saccharin	4	8	9	7	8	9	0
Aspartame	4	10	6	9	7	7	2
<i>High calorie taste</i>							
Sucrose	3	3	2	20	10	5	2
Fructose	1	3	4	10	11	9	7
Saccharin	0	4	6	15	9	8	3
Aspartame	2	3	5	15	6	12	2
<i>Liked the sweetener</i>							
Sucrose	7	1	9	11	4	9	4
Fructose	6	11	4	13	5	3	3
Saccharin	7	3	6	10	9	10	0
Aspartame	5	3	7	8	11	6	5

Note: Each row contains 45 judgments, 15 of which were respondent duplicates. For sweetness, artificial taste and high calorie taste, the level of the attribute decreases as the scale increases from 1 to 7. For aftertaste and liked the sweetener, the level of the attribute increases as the scale increases from 1 to 7.

Table 3: Number of respondents within each rating difference category for each of the five sweetener attributes

Sweetener attribute	Rating difference													Total
	–6	–5	–4	–3	–2	–1	0	+1	+2	+3	+4	+5	+6	
	Number of respondents													
Sweetness	0	0	1	5	3	8	10	12	8	8	2	1	2	60
Aftertaste	1	3	7	7	3	8	18	8	4	1	0	0	0	60
Artificial taste	1	1	3	8	1	14	13	7	5	2	3	2	0	60
High calorie taste	0	0	1	4	6	8	23	8	4	4	1	0	1	60
Liked the sweetener	0	1	2	8	4	4	17	7	5	5	3	3	1	60

Note: Rating difference is the first minus second judgment of the same product. For sweetness, artificial taste and high calorie taste, a positive rating indicates that the sweetener was judged to have more of the attribute (for example, more sweet tasting) in the second judgment than in the first. For aftertaste and liked the sweetener, a positive rating indicates that the sweetener was judged to have less of the attribute (for example, less aftertaste) in the second judgment than in the first.

across the five attributes are presented in Table 3. Depending on the attribute, between 10 and 23 of the 60 respondents had a rating difference of zero or were consistent in their ratings of the identical product. For each attribute, approximately 50 per cent of the respondents

were within 1 point of their first rating of the product; 80 per cent of the respondents were within 3 points of their first rating. A few judgments varied by as many as 6 points. Considering the attributes determined to be significant via the switchback design, sweetness

and aftertaste, only 10 and 18 of the 60 respondents, respectively, were consistent in their judgment of the same sweetener. The inconsistency of respondent judgments may be because of context effects or respondent ability.

The direction of the respondent judgment inconsistencies varied by attribute. The inconsistencies associated with the two attributes determined to be statistically significant via the switchback design, sweetness and aftertaste, were of particular interest. For sweetness, respondents tended to judge the same product as sweeter on the second tasting. For aftertaste, respondents tended to judge the same product as having a stronger aftertaste on the second tasting. Again, this may have been indicative of context effects, such as order or fatigue, as well as respondent ability.

Context effects

Context effects were analysed for the two significant attributes, *sweetness and aftertaste*. The mean difference ratings for sweetness for each switchback triad are presented in Table 4. For sweetness, it appeared that when a stronger sweetener, sucrose or saccharin, surrounded a weaker sweetener, fructose, the stronger sweetener was rated as more sweet on the second tasting. However, when a weaker sweetener, fructose or sucrose, surrounded a stronger sweetener, sucrose, saccharin or aspartame, the weaker sweetener was judged as less sweet on the second tasting. For sweetness, it appeared that the context of the judgment was important. For aftertaste, no significant differences were found among the mean difference ratings for the switchback triads. Therefore, it must be assumed that in this study no context effects existed for aftertaste.

Comparison with one-way analysis of variance

To determine the impact of respondent reliability and context effects, responses to the first product tested by each respondent were analyzed as a one-way analysis of variance. The one-way analysis of variance results were then compared to the switchback results. With one-way analysis of

Table 4: Mean difference ratings for sweetness for the 12 switchback triads

Switchback triad			Sweetness
1	2	3	Mean difference rating
Sucrose	Fructose	Sucrose	3.0 ^a
Saccharin	Fructose	Saccharin	2.6 ^{ab}
Aspartame	Fructose	Aspartame	1.4 ^{abc}
Saccharin	Sucrose	Saccharin	1.4 ^{abc}
Fructose	Aspartame	Fructose	1.0 ^{abc}
Aspartame	Sucrose	Aspartame	0.6 ^{abc}
Aspartame	Saccharin	Aspartame	0.6 ^{abc}
Saccharin	Aspartame	Saccharin	0.4 ^{abc}
Sucrose	Saccharin	Sucrose	0.2 ^{abc}
Fructose	Sucrose	Fructose	-0.2 ^{bc}
Sucrose	Aspartame	Sucrose	-1.0 ^c
Fructose	Saccharin	Fructose	-1.2 ^c

Note: ^{a,b,c}Means within column with unlike letters are significantly different at $P < 0.05$; $n = 5$ for each mean. A large, positive mean indicates that the identical sweetener was judged as more sweet on the second rating than on the first.

variance, no significant differences were found among the four sweeteners on any of the five attributes. However, the switchback design identified differences among the sweeteners. Thus, in the presence of respondent reliability and context effects, the switchback design was more efficient than the monadic, one-way analysis of variance.

CONCLUSIONS AND IMPLICATIONS

According to Batsell and Wind,³ a respondent can be expected to evaluate only a limited number of products at one time. Additionally, any product-testing design should provide for explicit testing of the reliability of the data. If the order in which the products are evaluated is expected to affect results, then the order should be pre-specified and the order effects should be analysed explicitly. Further, it may be possible to capitalise on order effects because they may help to identify the most vulnerable brand, suggest a particular advertising sequence and so on. The switchback design directly addresses these issues.

With paired comparisons, respondents test product samples in a paired format and render a judgment. Repeated paired comparison or triangle tests can be used to investigate product differences, context effects and respondent

reliability, and to calculate a Bayes true score of ability to identify experts.^{2,13} However with repeated paired comparisons or triangle tests, as the number of products involved in a test increases, the number of samples any one respondent is asked to evaluate quickly escalates. Increasing the number of attributes to be tested further multiplies the complexity of both the repeated paired comparison and triangle tasks. While expert respondents may be capable of evaluating a large number of samples, volunteer respondents may not.

In comparison, the switchback design is much less complex than either the repeated paired comparison or triangle tests. For the respondent, the switchback design is a simple triangle test of two products; for the researcher, it is a test of multiple products over multiple dimensions that provides a direct measure of respondent rating consistency and permits examination of context effects. For some products, context effects may be the research focus. If this is the case, the use of expert respondents within the switchback design may provide a more precise measurement of context effects. However, the switchback design does not provide the necessary information to identify expert panelists as do the repeated paired comparison or triangle tests.

With paired comparison and triangle tests, each respondent tests each product; therefore, more information is obtained per respondent. With the switchback design, each respondent tests only two products; this requires more respondents to be used in order to obtain information about all products in the test. For example, the switchback design requires 12 respondents per replication for a four-product test and 30 respondents per replication for a six-product test. This may be potentially disadvantageous when the pool of respondents is limited. However if the respondent pool is not limited, researchers are able to administer the switchback protocol to more respondents within a given time frame than the more complex repeated paired comparison or triangle tests. Whether testing in a repeated format or with the switchback protocol, the amount and cost of product required may represent an additional concern. Depending on

the research study, the cost of both respondents and products may be significant factors to consider.

There are several other advantages to the switchback design. First, respondents have no prior information that differences exist between the products, and are not pre-disposed to guessing. This is a departure from traditional triangle tests, where respondents are provided information about the products and may try to guess the correct response. Second, if respondents are affected to differing degrees by context effects such as fatigue or aftertaste, the design controls for this source of variation. Third, the switchback design requires that respondents be randomly assigned to treatments. When volunteer respondents cannot be pre-screened for treatment assignments, this can be a distinct advantage. Fourth, if testing involves multiple days or locations, the switchback design accommodates block effects.

In conclusion, for tests with multiple products and attributes where context effects are either not anticipated or can be controlled, monadic tests may be effective. In tests with multiple products and attributes where context effects are anticipated, the choice of research design is more difficult. The switchback design, repeated paired comparisons and repeated triangle tests all merit consideration. The switchback design is recommended for product-testing situations in which the following conditions exist: (1) multiple products; (2) multiple dimensions; (3) untrained volunteer respondents; (4) large numbers of respondents available; (5) one or more testing sites or days needed to complete the study; (6) short time periods in which to process individual respondents; (7) test administration simplicity is important; (8) measurement of respondent reliability is desired; (9) measurement of context effects is desired; (10) large quantities of the test products are available; and (11) the cost of the test products is minimal.

RECOMMENDATIONS

In many marketing considerations, millions of dollars are at stake, as firms attempt the risky process of introducing new products and

modifying existing ones in a dynamically competitive environment. Of necessity, many companies engage in field experiments in which untrained volunteer respondents are used to assess products characteristics and preference. In such cases, a nearly unlimited supply of respondents may be available and tests can be conducted simultaneously in numerous locations over a rather extended time frame. Additionally, these circumstances require simplicity and speed in testing. In many cases, the costs of the test products are minimal compared with the investments required to introduce and distribute these products, and large quantities of these products are available. Thus, it would appear that the switchback design would be both warranted and preferred in many product evaluations. Any disadvantages associated with the number of respondents required and with the identification of identifying expert panelists would seemingly be subjugated by desires for enhanced accuracy and reliability in the research design process. In conclusion, the switchback design promises more accurate feedback regarding both product preference and discrimination, which should lead to better decisions as they relate to both new product development and product modification.

REFERENCES

- 1 Ghose, S. and Lowengart, O. (2001) Taste tests: Impacts of consumer perceptions and preferences on brand positioning strategies. *Journal of Targeting, Measurement and Analysis for Marketing* 1(10): 26–41.
- 2 Buchanan, B. and Henderson, P. W. (1992) Assessing the bias of preference, detection, and identification measures of discrimination ability in product design. *Marketing Science* 11(Winter): 64–75.
- 3 Batsell, R. R. and Wind, Y. (1980) Product testing: Current methods and needed developments. *Journal of the Market Research Society* 22(April): 115–139.
- 4 Dahlstrom, R., Nygaard, A. and Crosno, J. L. (2008) Strategic, metric, and methodological trends in marketing research and their implications for future theory and practice. *Journal of Marketing Theory and Practice* 16(2): 139–152.
- 5 Buchanan, B. and Morrison, D. G. (1984) Optimal design of parity tests. *Journal of Mathematical Psychology* 28(December): 453–466.
- 6 Buchanan, B. and Morrison, D. G. (1985) Measuring simple preferences: An approach to blind forced, choice product testing. *Marketing Science* 42(Spring): 93–109.
- 7 Greenhalgh, C. (1966) Some techniques and interesting results in discrimination testing. *Journal of the Market Research Society* 8(October): 215–235.
- 8 Hopkins, J. W. and Gridgeman, N. T. (1955) Comparative sensitivity of pair and triad flavor difference tests. *Biometrics* 11(March): 63–68.
- 9 Moskowitz, H. R., Jacobs, B. and Firtle, N. (1980) Discrimination testing and product decisions. *Journal of Marketing Research* 17(February): 84–90.
- 10 Gruber, A. and Lindberg, B. (1966) Sensitivity, reliability, and consumer taste testing. *Journal of Marketing Research* 3(August): 235–238.
- 11 Roper, B. (1969) Sensitivity, reliability, and consumer taste testing: Some 'rights' and 'wrongs'. *Journal of Marketing Research* 6(February): 102–105.
- 12 Givon, M. and Goldman, A. (1987) Perceptual and preferential discrimination abilities in taste tests. *Journal of Applied Psychology* 72(2): 301–306.
- 13 Buchanan, B., Givon, M. and Goldman, A. (1987) Measurement of discrimination ability in taste tests: An empirical investigation. *Journal of Marketing Research* 24(May): 154–163.
- 14 Givon, M. (1989) Taste tests: Changing the rules to improve the game. *Marketing Science* 8(Summer): 281–290.
- 15 Morrison, D. G. (1981) Triangle taste tests: Are the subjects who respond correctly lucky or good? *Journal of Marketing* 45(Summer): 111–119.
- 16 Day, R. L. (1969) Position bias in paired product tests. *Journal of Marketing Research* 6(1): 98–100.
- 17 Welch, J. L. and Swift, C. O. (1992) Question order effects in taste testing of beverages. *Journal of the Academy of Marketing Science* 20(Summer): 265–268.
- 18 Buchanan, B. (1987) A model for repeat trial product tests. *Psychometrika* 521(March): 61–78.
- 19 Anderson, V. L. and McLean, R. A. (1974) *Design of Experiments*. New York: Marcel Dekker.
- 20 Cochran, W. G. and Cox, G. M. (1957) *Experimental Designs*. New York: John Wiley & Sons.
- 21 O'Mahony, M. (1974) Taste adaptation: The case of the wandering zero. *Journal of Food Technology* 9: 1–12.
- 22 Lucas, H. L. (1956) Switchback trials for more than two treatments. *Journal of Dairy Science* 39(February): 146–154.
- 23 Keith, N. K. and Williams, G. D. (1986) A generalized SAS algorithm including error diagnostics for the switchback experimental design with three or more treatments. In: *Proceedings of the SAS Users Group International*, SAS Institute Inc., Cary, NC, USA, Vol. 11, pp. 661–663.