Product Testing with Consumers for Research Guidance

Louise S. Wu, editor

STP 1035

ASTM

STP 1035

Product Testing with Consumers for Research Guidance

Louise S. Wu, editor

ASTM
1916 Race Street
Philadelphia, PA 19103

Library of Congress Cataloging-in-Publication Data

Product testing with consumers for research guidance/Louise S. Wu, editor.
(STP; 1035)
"ASTM publication code number (PCN) 04-010350-36"—t.p. verso.
Papers presented at a symposium held in Baltimore, MD on 17 May 1988 and sponsored by ASTM Committee E-18 on Sensory Evaluation of Materials and Products . . . [et al.]
Includes bibliographies and indexes.
ISBN 0-8031-1256-4
1. New products—Congresses. 2. Commercial products—Testing—Congresses. 3. Marketing research—Congresses. 4. Consumers' preferences—Congresses. I. Wu, Louise S. II. ASTM Committee E-18 on Sensory Evaluation of Materials and Products. III. Series: ASTM special technical publication; 1035.
HF5415.153.P76 1989
658.8′ 35—dc20 89-6901
CIP

Copyright © by American Society for Testing and Materials 1989

NOTE

The Society is not responsible, as a body, for the statements and opinions advanced in this publication.

Peer Review Policy

Each paper published in this volume was evaluated by three peer reviewers. The authors addressed all of the reviewers' comments to the satisfaction of both the technical editor(s) and the ASTM Committee on Publications.

The quality of the papers in this publication reflects not only the obvious efforts of the authors and the technical editor(s), but also the work of these peer reviewers. The ASTM Committee on Publications acknowledges with appreciation their dedication and contribution of time and effort on behalf of ASTM.

Printed in Chelsea, MI
July 1989

Foreword

This publication, *Product Testing with Consumers for Research Guidance, ASTM STP 1035,* contains papers presented at the symposium of the same name held in Baltimore, MD on 17 May 1988. The symposium was sponsored by ASTM Committee E-18 on Sensory Evaluation of Materials and Products, the Product Development Management Association, the American Marketing Association, and the American Statistical Association (Committee on National and International Statistical Standards). Louise S. Wu, Philip Morris USA, presided as symposium chairman and was editor of this publication.

Contents

此为试读,需要完整PDF请访问: www.ertongbook.com

Overview

The goal of this symposium publication is to facilitate discussion of approaches for the use of consumers in research guidance testing. Research guidance testing can be defined as the consumer testing that Product Development uses in designing modified, improved, or new products that consumers want and will buy when introduced. It is to the product developers what marketing research is to the marketers when trying to develop pack graphics and advertising. The consumers of a product need to be considered, in a carefully planned and systematic manner, at different stages of the product development process. What is presented in this STP will give some guidance as to how consumers can be included in the development process.

The first section, "Approaches to Research Guidance Testing," gives examples of successful development programs. The theme of these papers is involving the consumer, as the user of the product, in the evaluation of prototypes and directions for prototype improvement. This is an alternate to prototype selection by internal judgement. If the judgement prototype does not perform as expected, there will be no information on how to improve the prototype. By involving the consumers in product development, a final prototype that meets the needs and expectations of the consumer will be developed. Hlavacek and Finn use breakfast cereals and Baxter cites salad dressings as examples of products successfully developed with research guidance testing with consumers at different points in the development process. In the paper by Wolter et al., the consumer of the product (clothing labels) is industrial rather than individuals; however, many of the principles are the same.

The second section, "Processes for Research Guidance Testing," goes into more detail about different types of techniques that can be used. Carr discusses the use of designed experiments that include "response surface" to develop the optimum product. Cooper et al. discuss the concept of "ideal" points for product characteristics and how they can be used to segment the consumer population. Moskowitz and Jacobs describe the use of current products on the market as consumer benchmarks and how a new product can be designed against these benchmarks.

The final section is "Qualitative Research Guidance Testing." Over the past few years more and more product development groups are finding it useful to listen to consumers talk about their products to generate new ideas and to improve prototypes. Hayes presents a review of focus groups, including their uses and abuses, as well as what to consider in setting up a qualitative program. Younkin discusses the use of "kid peer leaders" in the development of products at American Chicle.

Overall, I believe the information presented is useful and will at least make the reader stop and think about how the product development process is carried out and how consumers can be used to make the process more successful.

Louise S. Wu

Philip Morris Research Center, Richmond, VA; symposium chairman and editor

Approaches to Research Guidance Testing

Douglas Hlavacek[1] *and John P. Finn*[1]

Hitting the Target More Frequently: A Systematic Approach to Research-Guidance Tests

REFERENCE: Hlavacek, D. and Finn, J. P., **"Hitting the Target More Frequently: A Systematic Approach to Research-Guidance Tests,"** *Product Testing with Consumers for Research Guidance, ASTM STP 1035,* L. S. Wu, Ed., American Society for Testing and Materials, Philadelphia, 1989, pp. 5-11.

ABSTRACT: One pitfall in product development is a tendency for developers to expend significant amounts of time, effort, and money on the technical aspects of the project, but leave to informal evaluation the selection of products that will be evaluated in business decision-making marketing research tests. The poor record of success of such an approach necessitates the development of a better system. Research-guidance testing has two primary goals, to provide (1) information concerning how acceptable prototypes are, and (2) direction concerning "next steps." A systematic approach is needed that provides developers with frequent, inexpensive, quickly executed, and accurate (that is, predictive of business decision-making test criteria) evaluations of prototypes. A rationale and program for obtaining consumer feedback and of integrating this information in a systematic fashion with product development from early concept/product screening stages through early development to the product-optimization stage are described, along with several examples of how the system operates.

KEY WORDS: research-guidance, designed experiments, consumer feedback, product development, early concept/product screening, early development, product optimization

Current State of Affairs in Research Guidance Testing (RGT)

The impetus for developing new products, and for developing quality improvements, cost reductions, or line extensions for existing products, can come from a variety of sources both within and outside a corporation (for example, new processing capabilities, competitive environment, suppliers' costs, marketing research forecasts, etc.). What is critically important is that a system exists that can guide product groups from the "ideas" stages through the product evaluation stages quickly, inexpensively, exhaustively, and accurately. It would seem obvious that the consumers' perspective would be routinely and importantly integrated into all stages of product development; surprisingly, this is often not the case. Product prototypes are often developed more with an "eye" toward costs, manufacturing potential, and what we think consumers want, rather than with an "eye" toward maximizing consumer acceptance.

A frequent consequence of this approach is that only after several iterations of development are products first evaluated by consumers. These prototypes often differ from one another in some nonsystematic manner. For instance, cereal prototypes may differ not only in type of base grain (for example, oat, corn, wheat) and sweetener system used (for example, sucrose, brown sugar, and honey), but also in the amounts of sweetener used with different bases. One is then left with the bewildering task of asking consumers to decide among alternatives that are as

[1]Technical Center and Consumer Response & Information Center, General Foods USA, White Plains, NY 10626.

different from one another as "apples and oranges." The "best" candidate(s) resulting from such a "system" is (are) then submitted for evaluation in a marketing-research test, often with not-so-surprising negative results.

This is an inefficient system at best for all of the obvious reasons, but what sometimes follows is even worse. If the overall results of the test do not support the business proposition, product developers will often look to the internal measures of product attributes on the test as a guide for what needs "fixing" in order to make the prototypes achieve higher acceptance on subsequent testing. Given the haphazard decision-making process that went into the selection of the original products for testing, and the subsequent guessing as to what factor (for example, ingredient, process) "caused" the poor evaluation, this too is bound for failure more often than for success.

There is reasonable evidence that people, despite their own claims to the contrary, cannot accurately report on their own internal "mental" states [*1*]. That is, asking a person to answer some 40 to 60 questions about everything from how "crunchy," "sweet," "salty," "moist," and "spicy" some food is, to "how much fun," "good as a snack," "something the whole family would enjoy," and then whether they prefer the product they tried this week more or less than the one they tried last week, may be neither realistic nor valid as a technique.

A Model for Research-Guidance Testing

The overall objective of research-guidance testing (RGT) is to increase the likelihood of identifying product(s) that, in the short term, will be "successful" in marketing-research testing, and, in the long term, will be "successful" commercial products.

In order to accomplish this objective, any RGT should be directed toward the following two goals:

1. Describing where we are with respect to how acceptable the prototypes are compared to other products drawn from a similar part of the market structure.
2. Pointing to where we can go from here in terms of what actions must be taken in order to increase some measure of acceptability (for example, preferences, overall liking).

In the best of all possible worlds, RGT is at least a three-stage process, each stage designed to narrow the range of potential products that are consistent with the design criteria: (I) the early concept/product screening; (II) the early development; and (III) the product optimization. Embedded within each stage is the concept that experimental design (See Refs *2* and *3* for examples), by its very nature a systematic approach, offers a better alternative. The importance of factors affecting consumer acceptance can then be determined with quantifiable precision, and complex effects can be efficiently unravelled.

I Early Concept/Product Screening

In this first stage, the outcome is to develop as wide a knowledge base about the product or product idea as possible. The strategy is to develop and evaluate the greatest number of variables and levels within these variables that fall within the range of parameters of the project. The outcome is to select for further development only those which demonstrate some specified level of consumer acceptability. This level, for example, may be based upon the performance of control products drawn from within some selected market segment.

The most critical aspect of this process is to choose those products that have the most leverageable effect on consumer acceptance. One hurdle that has to be overcome is the "mind-set" of product developers who have worked on the same product/type of product for a long period of time: the longer they have worked on it the shorter the list of variables they are willing

此为试读，需要完整PDF请访问： www.ertongbook.com

to consider. The result of this is that potential prototypes may be rejected based upon incomplete/incorrect information.

Given that we can identify a very large number of variables for a product, we need a system by which we can reduce these to some reasonable number. Screening-test designs, such as fractional factorials, are well suited at this stage to help identify those variables which most contribute to consumer acceptance.

This is a difficult step to accomplish because product developers sometimes need to expend significant resources to develop product prototypes they know will never be commercialized. They must be convinced to do this because the knowledge gained from these evaluations will make decision making easier as the development cycle progresses.

II Early Development

By this stage, we have already identified a reduced number of variables that meet some defined consumer-acceptance criterion. What we now need to understand is not only how important each variable is individually, but also how important combinations of these are in terms of acceptance.

Experiments using factorial designs are most appropriate at this stage. There are at least three advantages in doing this. One advantage of designed experiments lies in the simplicity of analysis through, for example, simple graphical methods. Secondly, we can reduce the amount of time the consumer spends filling out a questionnaire by limiting questions only to those which specifically address the experimental hypotheses. Thirdly, we can gain an understanding of how consumers translate our systematic product changes into their vocabulary. For example, is a change in perceived "strength of flavor" related to a change in the level of some flavor system or is it related, perhaps, to a change in sweetener level?

The next hurdle is that of specifying the range or limit for each of the quantitative variables identified. This can be an especially difficult task for an established product: after working for several years on the same product, developers often become sensitized to "microscopic" differences on the bench top which are transparent to the consumer. A good starting point, then, is to define each variable's limits based upon factors such as product stability, functionality, or manufacturability, for example. Whether these limits result in useful (that is, able to be detected) levels of variables can be verified through discrimination testing (for example, same-difference, triangle, etc.) prior to the formal RGT.

III Product Optimization

In this final stage, the "best" candidate(s) is (are) selected for the business decision-making marketing-research tests based upon a substantial convergence of information from the previous work. No longer is the major concern that of identifying which variables are important, but rather in "honing in" on the appropriate levels of the variables. Systematic research is now carried out over a very restricted range of variables and levels. Common designs used are multilevel factorials, central composite, or Box-Behnken designs.

At the end of this development cycle, developers should not only be instrumental in selecting which products will be tested in the business decision-making test, but also be in a position to influence how the test will be conducted and how it will be evaluated. In other words, as a result of these evaluations product developers are now in a good position to be involved in the design of the test and the setting of business risk profiles.

New Opportunities for Research Guidance Tests

Clearly, costs and time factors in RGTs are influenced by (*a*) the type of test procedure used, (*b*) the number of products sampled by each respondent, (*c*) the nature of the products to be evaluated, (*d*) the size and availability of the target group, and (*e*) the particular suppliers used. Of these factors, identifying the appropriate target group may be the most expensive and time-consuming.

We have used two other cost- and time-effective consumer resources, particularly during early concept/product screening. These include using company employees and consumers who have contacted the Consumer Response and Information Center of the company with a complaint/inquiry/testimonial through the toll-free telephone number or by mail.

Sensory-evaluation groups in companies have for years depended upon employee pools to carry out a large amount of needed discrimination and descriptive testing. For RGTs we extended this practice by recruiting employees in the same fashion as outside suppliers would. That is, they are recruited as users of certain products or categories of products and we carry out the full range of consumer testing (for example, central location test (CLT), home-use test (HUT), focus groups, using repeated-measures or randomized block designs, etc.). With respect to consumers who contact the Consumer Affairs Department, we have asked for their voluntary participation as a known user of certain products or categories of products and also carry out the full range of consumer testing. We realize the statistical limitations of these "not-so-random" samples. However, use of these pools of consumers is preferable to the "benchtop" decision-making often used by product developers during the development cycle.

Examples of Research Guidance Tests

1. Children's Low-Sugar Cereal

It was desired to develop a new ready-to-eat kids cereal with reduced sugar level. Early concept/product screening work suggested that the primary attributes to consider were cereal-base formula and sweetener level. It was additionally suspected that sweetener level might affect consumer acceptance in a nonlinear manner. Based upon this information, two different base formulas (control, new) and three equally spaced sweetener levels (labeled "low," "medium," and "high") were selected, forming a 2 × 3 factorial design consisting of six distinct prototypes.

Procedures—Presweetened cereal users ($N = 330$) aged five to twelve years were recruited for the study using an external supplier. Tasting fatigue and boredom were kept to a minimum by having each child evaluate only two of the six prototypes in a home-use test format. Using a multiple paired-comparison design, each of the six prototypes was paired with every other prototype (yielding 15 pairings ignoring order of presentation). Each of the 15 pairs was presented to groups of approximately 22 children. The questionnaire was purposely kept short and included a "like-a-lot" question on the first prototype tasted and an overall preference question after the second prototype.

Results—It was recognized some 30 years ago that efficiencies would be gained if each prototype were compared to every other prototype. Using the multiple paired comparisons (MPC) model, it is possible to realize a desired effective precision through the use of fewer judgments per pair [*4,5*]. This model has since been extended to include situations where a factorial design is imbedded among the prototypes [*6,7*].

Preference indices were calculated for all six prototypes using the actual observed preferences from each of the 15 pairings. These indices are relative measures of preference. For example, if the six prototypes were equally preferred, the indices would all equal 1/6 (1/number of products = expected index). Analysis of the preference indices confirmed what was apparent upon a visual examination of the data (see Fig. 1). A significant increasing linear trend in preference was detected among sweetener levels, and a marginally significant cereal base by sweetener-level

ESTIMATED PREFERENCE INDICES

SWEETENER LEVEL	CEREAL BASE CONTROL	CEREAL BASE NEW
LOW	.070	.066
MEDIUM	.135	.162
HIGH	.225	.342

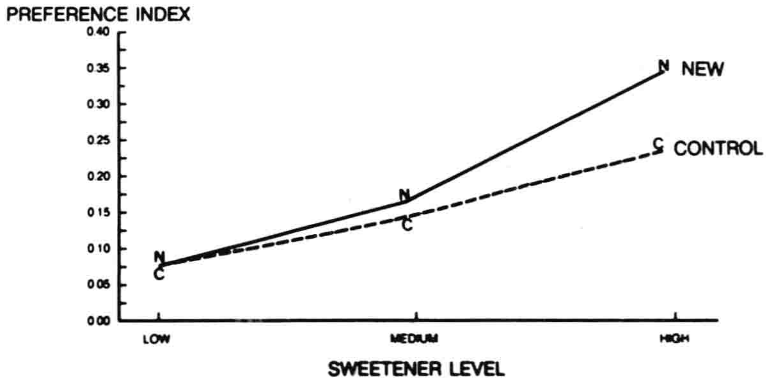

FIG. 1—*Estimated preference indices as a function of sweetener-level for "new" cereal-base product and for "control" base.*

interaction was found suggesting that the higher sweetener levels benefit the new cereal base more than the control base.

Discussion—A clearer understanding of the motivating factors behind children's preference was obtained through a systematic manipulation of two product-formulation variables. Subsequent broadscale marketing-research testing supported the business proposition; the RGT saved not only time and money but also identified a commercializable product.

2. Adult Blended Cereal

Previous product screening and early development work identified an opportunity to create a new adult product through the blending of three cereal bases (A,B,C). Subject to some preliminary evaluations and cost constraints, it was decided that a blend of A = 30%, B = 35%, and C = 35% would yield a highly acceptable product. It was desired to determine what effect, if any, the current process variation would have on consumer acceptance. A review of manufacturing data generated by the blending operation identified normal operating limits for each component to be about ±20%.

Given the multiple constraints on the three component proportions, an extreme vertices design for a linear blending model was selected [*8*]. Thirteen prototypes were developed (see Fig. 2). Design point "1" corresponds to the grand centroid, while points "2" through "7" correspond to the extreme vertices. Six additional points ("8" to "13"), not formally part of the design, were included at the developer's request.

POINT	A	B	C
1	30%	35%	35%
2	10	55	35
3	30	55	15
4	50	35	15
5	50	15	35
6	30	15	55
7	10	35	55
8	35	35	30
9	35	30	35
10	30	40	30
11	25	40	35
12	20	35	45
13	40	35	25

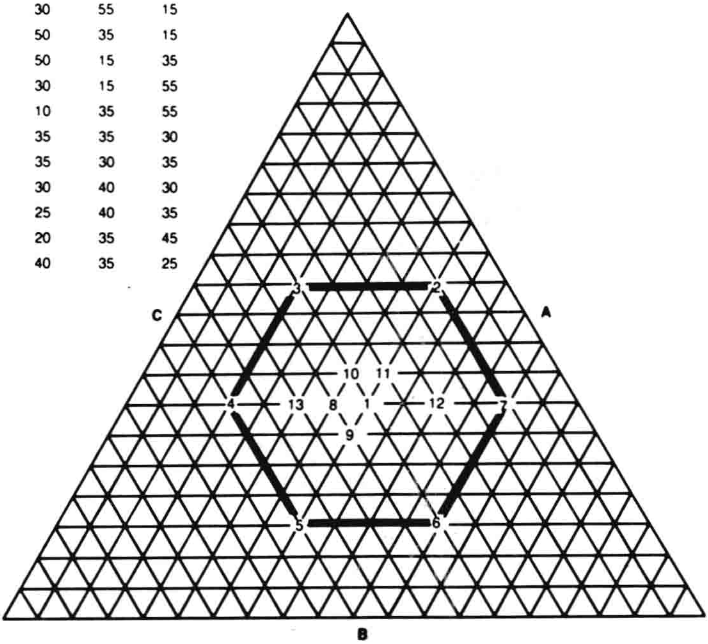

FIG. 2—*Schematic of mixture-design for three cereal components. Each point represents the proportion of Bases A, B, C.*

Procedures—Employees ($N = 104$), screened for usage of a class of adult cereals, were each asked to evaluate three of the thirteen prototypes for overall acceptability on a nine-point scale (1 = dislike extremely to 9 = like extremely) using a CLT format. Since each consumer did not evaluate all thirteen prototypes, an extended incomplete Latin-square design was used [*3*]. This plan enables each prototype to be presented with every other prototype an equal number of times. In addition, each prototype was presented in every tasting position (first, second, third) an equal number of times, with a few minutes elapsing between successive samples.

Results—After adjusting the prototype means for block effects, a linear blending model was fit to the overall acceptability of the seven design points. The analysis indicated that the surface above the constrained mixture region was a level plane whose height was the same for all points. A visual inspection of the adjusted prototype means supports the notion that consumer acceptance was relatively flat over the normal operating limits of the process (see Fig. 3).

Discussion—Experimental design proved to be an effective way of determining the compatibility of current process variation with consumer acceptance. In addition to determining the location of the optimum product (that is, where to set the production targets), we are just beginning to develop more realistic in-plant specifications which integrate consumers' input with technical research and manufacturing requirements.

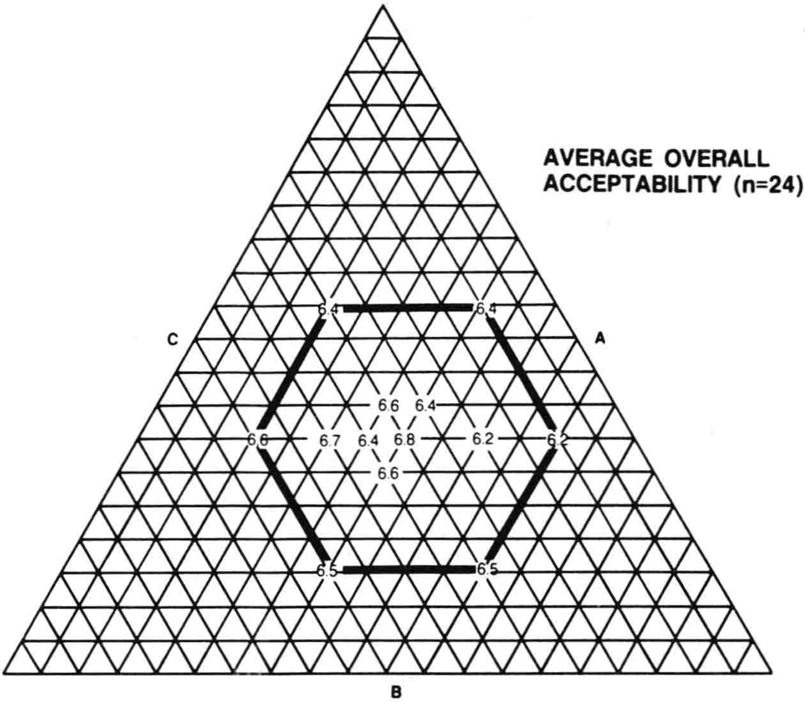

FIG. 3—*Overall acceptability ratings (9-point scale) of 13 samples schematized in Fig. 2. Each point is based upon 24 judgments.*

Acknowledgments

Special thanks to Bob Griffin (manager of Sensory and Statistical Sciences) and Dick Fizzell (Post Cereals' research manager) for their help and assistance. We would also like to thank the anonymous reviewers for their comments.

References

[*1*] Nisbett, R. E. and Wilson, T. D., "Telling More Than We Know: Verbal Reports on Mental Processes," Psychological Review, Vol. 84, No. 3, May 1977, pp. 231-259.
[*2*] Box, G. E. P., Hunter, W. G., and Hunter, J. S., *Statistics for Experimenters,* Wiley, New York, 1978.
[*3*] Cochran, W. G. and Cox, G. M., *Experimental Designs,* Wiley, New York, 1957, pp. 376-544.
[*4*] Terry, M., Bradley, R., and Davis, L., "New Designs and Techniques for Organoleptic Testing," *Food Technology,* Vol. 6, No. 2, July 1952, pp. 250-254.
[*5*] Bradley, R., "Incomplete Block Rank Analysis: On the Appropriateness of the Model for Method of Paired Comparisons," *Biometrics,* Vol. 10, No. 3, September 1954, pp. 375-390.
[*6*] Abelson, R. and Bradley, R., "A 2 × 2 Factorial with Paired Comparisons," *Biometrics,* Vol. 10, No. 4, December 1954, pp. 487-502.
[*7*] Bradley, R. and El-Helbawy, A., "Treatment Contrasts in Paired Comparisons—Basic Procedures with Applications to Factorials," Technical Report 348, Florida State University, Tallahassee, FL, July 1975.
[*8*] Cornell, J., *Experiments with Mixtures,* Wiley, New York, 1981.

Nancy E. Baxter[1]

Research Guidance: Not Giving It Your "Best Shot"

REFERENCE: Baxter, N. E., **"Research Guidance: Not Giving It Your 'Best Shot,'"** *Product Testing with Consumers for Research Guidance, ASTM STP 1035,* L. S. Wu, Ed., American Society for Testing and Materials, Philadelphia, 1989, pp. 10–22.

ABSTRACT: Research guidance testing can use experimental design and statistical analysis to assure Product Development of success with the consumers. There are two testing approaches for developing a product: the "best shot" approach and the multiple-sample experimental design approach. The advantages of the multiple-sample design approach and its benefits are presented using a series of research guidance tests on salad dressing as examples. The use of sample presentation designs and experimental designs (screening designs, optimizing designs, and factorial designs) enables us to obtain the maximum amount of information from our tests. In addition to the discussion of these designs, data analysis techniques for both normal theory and nonparametric procedures are considered. While multiple-sample experimental designs generally require more planning and initial capital investment than the "best shot" approach, the information gained and total cost savings are greater.

KEY WORDS: experimental design, balanced incomplete block design, data transformation, hedonic scale, preference, mixture experiment, split-plot design, screening design, optimization, factorial design, regression analyses, nonparametric analyses

Product development can be approached in two ways: the "best shot" approach (two sample), where our best guess is tested against the standard or the competition; or the experimental design approach (multiple-sample), where what is tested is planned by introducing systematic variation into the product. While the design approach requires more planning and more initial capital investment, the rewards are substantially greater than the "best shot" approach. This paper describes a case study for product improvement and cost reduction that demonstrates the advantages of using multiple-sample experimental designs. The feasibility of conducting designed experiments within a product development environment will be shown.

In this case study, a progressive series of research guidance tests was used to develop an improved and cost-reduced salad dressing. Seven tests were conducted over a period of twenty months. The design of each test was based on results of the previous test. At the conclusion of the tests, product formula changes were implemented.

As each test is presented, the issues concerning the experimental design and the analysis of the data will be discussed. The product improvements may yield an increase in profits 15 to 20 times the cost of the tests over the next five years.

Procedures

Throughout the series of tests, sample presentation designs and statistical analysis procedures were similar. Details of these procedures will be presented here.

[1]Statistician, Thomas J. Lipton, Inc., Englewood Cliffs, NJ 07632.

此为试读，需要完整PDF请访问：www.ertongbook.com