On the Conduct of In-Store Field Experiments

Contributed by Ronald C. Curhan
Department of Marketing
Boston University
Boston, Massachusetts

Provides specific information which needs to be considered by those responsible for the approval, design, conduct and analysis of in-store experiments.

Researchers frequently seek the cooperation of retailers to conduct experimental research under actual, on-going operating conditions. In particular, researchers seek to manipulate one or more merchandising variables such as retail price, media advertising, type of display, point-of-purchase advertising material, packaging or merchandise mix, among others; and to measure the resultant impact on sales and profit.

Apparently only rarely do retail organizations undertake such research for their own benefit, except on an informal, ad hoc basis. Experiments utilizing formal experimental designs generally are conceived by academic or manufacturer-sponsored researchers who seek answers to questions of only marginal interest to retailers. For example, an experiment to compare the appeal of alternate package designs may be critical to a particular manufacturer, but promise comparatively little in the way of net payback for efforts expended by a retailer. Yet, researchers continue naively to request retailers to conduct such experiments. Increasingly, retailers are refusing to undertake research which primarily benefits others,[1] or, alternatively, are charging for the use of their store "laboratories" (see, for example, [5]). Clearly, researchers should appreciate that, in the absence of direct cash payments, anticipation of useful findings is the key to gaining retailer cooperation.[2]

Because opportunities to conduct experimental research are scarce and expensive, researchers must avoid wasting them on fishing expeditions and trivial replications of past studies. This is not to say that there is no place for true exploratory research or replicative studies, but rather to argue against undertaking projects which are not carefully drafted and hold the promise of yielding important data. It seems especially critical that experiments test, well-articulated hypotheses which, in turn, are based on sound theory and extend the results of previous research.

Let us assume that a worthy research question has been identified, and that a retail organization, say a supermarket chain, has agreed to cooperate by conducting an experiment. What, then, are the major operational issues of concern to the researcher?

ORGANIZATIONAL CONSIDERATIONS

The successful conduct of in-store research is dependent upon organizational no less than experimental design. The strong support of top management is necessary but not sufficient to insure successful execution of an experiment. Most likely the researcher will have negotiated for approval of the project with high-level managers of the retail

organization. While they, in turn, may have consulted with subordinates in the process of making their decision, undoubtably the researcher will have to depend upon many who were not involved in the decision process for execution of the experiment. It is absolutely essential to the success of any project that a sense of cooperation and commitment be fostered among all involved parties--headquarters executives, merchandisers, store supervisory and operating personnel, and, very often, advertising, warehousing, buying and auditing personnel as well. The wise researcher will see that personnel at all levels of the organization are properly motivated to cooperate--that is, he will see to it that top management makes known throughout the organization its commitment to the successful conduct of the experiment.

Communication and cooperation with and between line and staff, headquarters and stores, managers and clerks are critical to the success of any experiment, but especially so for multi-store projects. The researcher should initiate a genuine dialogue with those involved at a very early stage of his project. Not only can the inputs of those familiar with company procedures materially improve design and execution of the experiment, but their involvement in a meaningful way may result in personal commitments to the success of the project--a critical factor when unexpected operational difficulties have to be overcome.

Special note needs to be taken of the resistance which may be expected from operating managers who frequently perceive research as interfering with their job, increasing costs, and thus jeopardizing profits--the standard by which they are judged. Because experiments frequently do impinge on normal operations, management needs to make explicit its cognizance of this condition.

Successful execution of experimental research frequently is contingent upon retailer agreement to provide sufficient test stores, to allow necessary in-store and/or extra-store promotional manipulation and, in general, to provide a positive climate for the research. Usually, compromises will be necessary. Rarely, if ever, will a retailer agree to undertake or be able to accomplish an experiment exactly as proposed. Opportunities for trade-offs usually exist between, say, the number of test stores and the length of an experiment, or the number of levels at which a given variable is tested. It is crucial that the researcher understand which requirements are essential to the success of the experiment, and which can be finished safely. Much as one might want to pursue a project, it is foolish to proceed if essential design requirements cannot be guaranteed.

TEST STORE SELECTION

In addition to usual experimental design consideration, formulation of in-store field experiments usually necessitates researcher attention to test store selection. Four criteria generally are at issue: unit size, geographic location, quality of store personnel and characteristics of clientele.

Test Store Unit Size

There is some evidence that experimental results differ for large and small stores, or, more accurately, stores with high versus low sales volume [6], although this may be a function of our inability to measure accurately the changes in sales attributable to test variables in low volume stores [3]. Nevertheless, large units are to be preferred because a.) their greater activity essentially increases sample size, and b.) if results lead to modification of current practices, these changes are most likely to be implemented in larger

stores. For supermarkets, large stores may be defined as those having sales of $100,000 or more per week. Unless store size is a test variable, every effort should be made to utilize stores of like size and sales volume.[3]

Interstore sales variability is such that store replications invariably are in order. As a practical matter, at least four store replicates seem desirable and two would seem to be the absolute minimum. Formal experimental designs should be used wherever possible to facilitate statistical analysis.

Geographic Location of Test Store

In addition to selecting stores reasonably accessible in terms of travel time from either the chain headquarters or the researcher's base of operations, it also is desirable to select stores within a single promotional zone and to select as test units those stores which have as their principal competitors stores of a single competing chain. This facilitates monitoring of competitive activity and accounting for its likely impact. Likewise, the number of supervisory zones should be minimized. These factors outweight considerations of geographic diversity. Indeed, as will become evident, I will argue for use of the most homogeneous store sample feasible.

Quality of Test Store Personnel

Often, during the process of test store selection, the researcher will hear from management "We'll use Store Number 81. The manager there is very cooperative." Fine, but I would argue that the planning and organization for the conduct of in-store experiments must be so all inclusive as to insure that the successful execution of the experiment is independent of the individual attitudes or attributes of particular store personnel.

Characteristics of Test Store Clientele

Ethnicity and social class of store clientele should be as homogeneous as possible across test stores, unless these factors are specific test variables.[4] Variations reduce the likelihood of attributing test results to the test variables. Moreover, given experimental results for "typical" stores, managers seem capable of doing a good job of extrapolating results to stores with different patronage characteristics.

TEST ITEM SELECTION

Researchers often are interested in the effect of treatment variables on classes of products although specific items are tested to represent these classes of products. When selecting test products, researchers often define their choice of test item as "coffee", or "two-pound Maxwell House ground coffee". Selection then is justified on the basis that the item is "representative" of goods in a merchandise category. However, characteristics of products within conventional supermarket merchandise categories frequently differ greatly. Consider, for example, packaged pudding mix and canned puddings (desserts), or two-ounce olive oil, 48-ounce Wesson Oil and three-pound Crisco (oils and shortenings).

Test products may better be chosen on the basis of a taxonomy incorporating such factors as whether an item is the leading brand, second brand or follower in its merchandise category; whether or not its sales pattern is seasonal; whether it has a high or low advertising-to-sales ratio; whether it is an impulse item or a staple; or, according to package size, stage in its product life cycle, or similar consideration. It is argued that test items chosen on the basis of these criteria will have the greatest generalizability thus facilitating extrapolation of results to other like

products which may or may not be in the same conventional supermarket merchandise category as the test item. In today's dynamic store environment, study of a specific brand/size item "representatively" chosen may be so situation specific as to have limited usefulness. Admittedly, little attention has been given to development of such a taxonomy and operational measures of many variables are lacking. For example, there is no satisfactory measure of impulse. Much work needs to be done in this area.

Most experiments involving groceries consider only warehoused dry grocery products. This is understandable given operational constraints, although I would argue strongly for the inclusion of so-called "store-door" delivered goods in grocery studies as these represent an appreciable percentage of supermarket sales. The exclusion of beverages, bakery products, cookies and crackers, snacks and other store-door items from the typical grocery product study would seem to severely limit the generalizability of the results. Although movement data for these products is less likely to be computerized, inclusion of store-door delivered products should not be an insurmountable problem.

Needless to say, experimental designs should provide for replication within product classifications.

LEVELS OF TEST VARIABLES

Researchers frequently test a given variable at "high" and "low" levels. The results of such dichotomous studies often are limited in operational usefulness, especially in the absence of good understanding of threshhold levels of effectiveness of merchandising and promotional variables. Therefore, it is desirable to test at several levels to define the functional response to a given variable. It may be argued that, given practical constraints as to the size of experiments, this objective should take precedence over design considerations to maximize, for example, the number of test items. In other words, it is probably better to generate more data points on a few items, than to have sparse data on many items.

It may be that two-stage testing is called for: a "rough cut" to identify important variables, and "fine tuning" to more precisely define their effects. Fractional factorial designs would seem to be particularly useful to this process, although they have been little used in store research [7, 4].

COMPLIANCE AND MEASUREMENT

Compliance with Experimental Design Requirements

Compliance involves the manipulation of test variables according to plan and, equally important, the nonmanipulation of control variables. It may be that the researcher should assist or, better yet, assume full responsibility for making price changes, building and removing displays, and performing other tasks called for by the research plan. At the very least, he should plan to monitor these activities to insure timely execution and sustained compliance.[5]

Store personnel commonly are concerned with what will happen to test merchandise remaining unsold at the completion of an experiment. A commitment to provide for return of any excess to the warehouse or for its transfer to other stores will go a long way toward insuring cooperation. If original shipping containers are saved, this should be no great problem. (Excess goods removed from display have an understandable propensity to again appear on display once the researcher has left a store, a

practice which can wreak havoc with post-test period measurement.)

Avoidance of intra-store competitive activity is best handled by censoring office-to-store promotional communications (i.e., "sales planners", etc.) before they reach test stores, although it is unlikely that the researcher will be allowed to veto competing promotions which involve media advertising. Substitute promotional instructions can then be given. Of course, the researcher is hardly likely to be given carte blanche to make these changes. He should, however, establish routine procedures whereby he is able to discuss his requirements with company personnel who have the authority to order necessary changes.

Measurement of Experimental Effects

The success of any experiment is dependent upon the collection of accurate movement data for test periods. The usual method of collecting such data for experiments which extend over any length of time is to analyze delivery records and to adjust recorded receipts according to changes between beginning and ending inventory levels. Manual inspection of delivery records is less desirable than capturing movement data from EDP records, although the latter may require special programming and computer runs. In either case, adjustments must be made for merchandise transfers, returns and special deliveries of goods.

Calculation of rates of movement from delivery records and physical inventories is greatly superior to imputed measures of rates of sale based only upon warehouse-to-store shipments. If beginning and ending physical audits are impossible, care should be taken to insure that test periods are of sufficient length to insure that period movements are large multiples of likely beginning-to-ending inventory variations [2].

Audits should be conducted either by the researcher or by other nonstore personnel. Where company employees must be used, consider employing clerks from other than the test stores. If test store employees must be used, it is best to assign them on overtime as it is dangerous to make them responsible for performing the auditing task as part of their regular duties because operating necessities inevitably will take priority over test responsibilities.

Audits should be conducted using specially prepared forms and timed just prior to receipt of replenishment stock deliveries if the experiment extends over more than one delivery interval. (Anyone who has tried to tally reserve stocks buried under a mountain of incoming goods will attest the desirability of this timing.) Special care should be taken to insure that shipments are attributed to correct periods. The dating of invoices may differ by days from the time of actual delivery.

Perhaps the best insurance of accurate data collection is the immediate processing of audit and delivery information. Apparent discrepancies may be resolved if they are noted quickly, but hardly ever at a later date. The same holds true for promotional compliance and other facets of the experimental design.

Successful measurement is greatly dependent upon adequate experimental controls. Supermarkets comprise a dynamic environment. Chain management actively encourages changes which, unfortunately, are particularly disruptive to the conduct of research. New item introductions, product discontinuances, merchandise re-sets, and price changes; as well as stock-outs, deal packs and package changes all are potentially damaging to the designed experiment. Typically the researcher tries to "freeze" these factors for the duration of the research, at

least within the merchandise categories under consideration. Unfortunately, this task is complicated by activity of identified and, worse, unidentified test item complements and substitutes, many of which are outside the test categories. Perhaps the best protection against contamination by uncontrolled environmental variables is the rapid execution of the experiment, although this complicates the determination of whether or not the test period is "typical" along other dimensions.

Automated Checkouts

Special mention needs to be made of the potential for improving in-store experiments which may be expected as the result of adoption of automated checkout systems by supermarkets [1]. These systems have the capability to accurately record rates of sale for test items and, indeed, for those utilizing lazer scanning techniques, to read Universal Product Code symbols for a large majority of products sold. Moreover, sales data can be collected for any time interval.6/

This capability may make feasible a new level of sophistication in in-store field experiments wherein test variables may be changed at frequent intervals. For example, the impact on sales of a particular point-of-purchase sales message might be compared to another message by alternating each message on an hourly basis in test stores.

Moreover, these systems may be the basis for organizing a new type of consumer panel. Panel members might identify themselves at the time of checkout and a complete record of their purchase could be captured. Individual customer orders might then be systematically analyzed to yield data, for example, on the incidence of joint purchase of items displayed in a complementary fashion. On a longer term basis, brand switching and other responses to promotional campaigns might be measured. The potential of these new systems to improve data collection is not to be underestimated, although the opportunity for "outside" researchers to benefit from their use is likely to be very limited, at least in the near future.

SUMMARY

The in-store field experiment is an important research tool. However, the maximum benefits of this type of research may not have been realized in many instances because of limited attention to operational considerations by researchers. This article has focused on certain practical aspects of experimental design and execution. It is hoped that attention to these details will improve the relationship between researchers and cooperating retail organizations, and that it will also improve the quality of resulting research.

FOOTNOTES

1/ An analogy may be drawn to manufacturer reluctance to answer questionnaire requests for information reported in [8].

2/ Information on the relationship of shelf space, retail pricing and point-of-purchase advertising to product rates of sale are examples of questions historically of interest to retailers.

3/ If units do differ, stratified sampling may be appropriate.

4/ Again, if there are variations stratified designs should be considered.

5/ Special care must be taken to guard against overzealous compliance with design requirements. If experimental

conditions are intended to be "representative" of usual conditions, then excessive attention to neatness and fullness of displays is likely to bias results!

6/ Comparable systems have been developed for department and discount stores and other retailers.

REFERENCES

(1) Bloom, Gordon F. and Ronald C. Curhan. "Technological Change in the Food Industry", Technology Review, 77 (December 1974), 20-29.

(2) Curhan, Ronald C. "The Relationship Between Shelf Space and Unit Sales in Supermarkets", Journal of Marketing Research, 9 (November 1972), 406-12.

(3) Curhan, Ronald C. "Shelf Space Allocation and Profit Maximization in Mass Retailing", Journal of Marketing, 37 (July 1973), 54-60.

(4) Curhan, Ronald C. "The Effect of Merchandising and Temporary Promotional Activities on the Sales of Fresh Fruits and Vegetables in Supermarkets", Journal of Marketing Research, 11 (August 1974), 286-94.

(5) Elson, Joel. "Food Fair Plan Offers Makers An 'Open House' for Research", Supermarket News, 23 (May 20, 1974), 6.

(6) Frank, Ronald E. and William F. Massy. "Shelf Position and Space Effects of Sales", Journal of Marketing Research, 7 (February 1970), 59-66.

(7) Holland, Charles W. and David W. Cravens. "Fractional Factorial Experimental Designs in Marketing Research", Journal of Marketing Research, 10 (August 1973), 270-6.

(8) Reuschling, Thomas L. and Michael J. Etzel. "The Disappearing Data Source", Business Horizons, (April 1973), 17-22.

******