MARKETING SCIENCE
Vol. 3, No. 3, Summer 1984
Printed in U.S.A.

AWARENESS FORECASTING MODELS—EDITORS' COMMENT

The preceding paper by Mahajan, Muller and Sharma provides what one rarely finds in the Marketing literature—an empirical comparison of several alternative models. While the authors' objective is modest (comparing a *single* relationship: the advertising-awareness submodel of several new product models on two common data sets), they are to be commended for doing this research. This kind of empirical work enhances our understanding of the strengths and weaknesses of alternative models and also indicates directions for future research. For instance, Mahajan et al. had to simplify the nature of the awareness model for their comparisons (see the following comments). Future comparisons will hopefully be able to retain some of these complexities and also be able to compare the trial and repeat purchase modules of the various models.

It is also interesting to read the commentaries on the Mahajan et al. paper by the authors of the five models that are compared. Such discussion and debate can only increase our understanding of the complex issues related to modeling new product introductions. We want to encourage more such comparisons and commentaries, and hope that the reader finds the entire package (the paper, the commentaries, and the authors' rejoinder) to be both interesting and stimulating.

COMMENT

MARKETING SCIENCE
Vol. 3, No. 3, Summer 1984
Printed in U.S.A.

JOSEPH BLACKBURN AND KEVIN CLANCY
Vanderbilt University
Boston University

We like the Mahajan, Muller and Sharma paper for several reasons. First, it represents one of the few attempts in the marketing literature to compare empirically alternative models of a marketing process.[1] Second, the results are favorable to our model, LITMUS, fitting the data as well as the four competing models. Finally, the paper convinces us that our recently introduced LITMUS II model[2] would perform even better, since it is superior to the original in every respect, including the forecasting of awareness.

Our appreciation of the paper and LITMUS's performance aside, several aspects of the paper are disturbing—an oversimplification of the awareness generating process; the questionable validity of the test data; a focus on fitting rather than prediction, which can mask the real difficulties in successfully predicting new product performance; finally, some misunderstandings concerning LITMUS (due, in part, to the fact that our papers on the model are not as detailed as they might be).

[1] Several major advertisers routinely run such tests using their own data.

[2] See Blackburn and Clancy, "LITMUS II: An Evolutionary Step in New Product Planning Models from Marketing Plan Evaluation to Marketing Plan Generation," Proceedings of TIMS Marketing Science Conference, Los Angeles, 1983.

0732-2399/84/0303/0198$01.25

The Oversimplification of the Awareness Generating Process

The forecasting of new product awareness is a difficult problem since awareness is a *complex* function of *many* variables including absolute levels of media exposure (often measured in terms of gross rating points), relative levels of media exposure (or "share of voice"), media type, media scheduling, advertising intrusiveness, promotion type and frequency, distribution, retention and other factors. Mahajan et al. drastically simplify the problem by comparing the models in terms of *only one variable*, gross rating points. Thus their analysis is not really a comparison of new product awareness forecasting models but, rather, a very simple comparison of the models' GRPs-to-awareness functions.[3]

Data of Questionable Value

A closer look at the data sets employed reveals an equally serious problem: Most of the data analyzed in this paper appear to be for established products rather than new products. Seven of the eight brands analyzed in the paper were provided by Golanty and Associates and these, in our view, appear to be for established, rather than new, products. There are three reasons for our concern. First, the spending patterns in each case are more typical of established products campaigns—there is little evidence of "front loading." Second, there is a strange insensitivity to massive doses of media exposure. Third, the most compelling reason for our doubts is the *extremely* elevated levels of awareness, starting high (53% to 81%) and ranging during the third period from 71% to 87%. For two brands, B_1 and B_2 in Table 6, small injections of GRPs, 330 and 430, respectively, are producing freakishly high first period awareness levels of 58% and 61%. New products often end the first year where these scores begin and only achieve the third period levels by the third year if ever. The authors indicate that the data came from an established product category. We also suspect that the data came from established products in that category (possible line extensions to products which are household names) or, at best, products which were "new" two or three years earlier.

Since no new product modelers with whom we are acquainted would use a basic *new product* GRP-to-awareness function to forecast awareness for an established product (they are different phenomena), we are left wondering about the value of this exercise and are puzzled that this evaluation appears to be based on data neither LITMUS nor any of the other models was designed to forecast.[4]

Fitting Rather than Prediction Oversimplifies the Problem

Here we will direct our comments to the objective of the GRP-to-awareness data analysis—using OLS regression to estimate model parameters in order to compare fitted forecasts with actual numbers. Admittedly, TRACKER and the early NEWS presentations highlight fitting real-world data to their models—taking, for example, early test market *A*wareness, *T*rial and *U*sage (ATU) data and using it to forecast year-end performance. However, this is not the way most new product models are used today.

[3] Since most of the models are descendants of NEWS, it is not surprising that their design and performance in terms of this overly simplified problem turn out to be quite similar. Although NEWS was published in 1982, TRACKER, LITMUS etc. are, in fact, descendants because NEWS was privately published, widely circulated and routinely employed from 1971 or earlier.

[4] In our experience, new campaigns for established products can be successfully forecast, substituting the criterion variable of brand awareness with a campaign tracer element (a copy/visual stimulus unique to a campaign such as a slogan, a spokesperson etc.) and by incorporating some of the additional factors (relative media exposure, media type, attention getting power etc.) cited earlier.

As discussed in our original paper on LITMUS (1980, 1982) managers use models at three different stages—in pre-laboratory test market simulation, in conjunction with a laboratory test market, and as an analytical tool superimposed on test market tracking data. Of approximately 300 LITMUS projections run during the past year, at least 85% were employed in the first two stages to forecast awareness through sales using a modest amount of data for the new product. The authors' approach is similar to taking early test market ATU data and using it to estimate model parameters in order to forecast year-end performance. This is a far easier task than most new product forecasters face because when data on the performance of a plan are available for the first six months, it takes neither a genius nor a complex model to forecast year-end with reasonable accuracy.

Prior to the real world test, when data on actual plan performance are nonexistent, a sophisticated model is needed to capture all the nuances of a marketing plan in order to forecast accurately. At this point, *no standardized procedure exists for fitting the values of* A_0, α, β, k, *etc*. Instead the quality of the forecast depends less on curve-fitting capabilities and more on the skill of the practitioner, the quality of marketing research, model structure and a sound historical data base.

In addition, there are important issues of statistical analysis which we will only list in passing: instability of OLS parameters based on samples of 5 and 3; measurement error in real world awareness estimates; and the artifactual discovery that initial awareness is the key parameter, a result we suspect follows from using awareness data for established products, data with incredibly high (for a new product) initial awareness levels.

Some Misunderstandings Concerning LITMUS

Mahajan, Muller and Sharma suggest that the forgetting effect in LITMUS does not depend on the awareness level achieved and do not use a forgetting coefficient in their curve-fitting tests of the LITMUS model. Forgetting is explicitly modeled in LITMUS (as in the NEWS model) using a retention coefficient, $r_c(i)$, to "denote the probability that a consumer in awareness state i will retain awareness in the succeeding purchase period," and, as such, depends directly on the specific awareness level. In fact, using a retention factor of 0.9 and a β value equal to that used by NEWS, LITMUS produces results equivalent to the NEWS predictions in Tables 3 and 6. The retention coefficient and an initial awareness factor are estimated as in NEWS using normative data, management judgement and/or primary research.

Far more important than forgetting or initial awareness are factors such as relative levels of media exposure (e.g. "share of voice"), media type, the vehicle within media type (e.g. "Hill Street Blues" vs. "Simon and Simon"), advertising intrusiveness, and couponing/sampling drop frequency and type. Unhappily for us, Mahajan et al. fail to discuss these factors. We feel that LITMUS comes closer to addressing satisfactorily all of these factors than the other models studied.

The authors suggest a need for further information on our data sources for LITMUS. Since LITMUS evolved from NEWS, which was routinely employed by BBDO as early as 1970, the cases around which NEWS was built and tested provided a foundation for LITMUS. Moreover, our 1980 and 1982 papers acknowledged our later relationship with the research firm of Yankelovich, Skelly and White, Inc. as a data source for early tests with 24 new products. Since 1982, 52 more validation studies have been completed. In addition, we have enjoyed access to the YSW library of new product simulation and tracking research to aid in model calibration and have profited from discussions with YSW executives and clients about what was right and what was wrong with the original LITMUS formulation.

Conclusion

After beginning our commentary by commending the authors for their pioneering effort, we moved on to criticize, constructively we hope, their study. In conclusion, we should candidly admit how difficult it is to evaluate marketing models in terms of predictive power. Our experience has shown that these models require vigilant intervention to insure that the data inputs and parameter estimates are the best possible before making any forecast.[5] It is extremely difficult for an outside observer to take a published marketing model of some complexity and use it to make the type of forecasts managers want and against which validation cases can be written. Only by reducing the model down to a simple base level—as was done by abstracting the awareness functions in this paper—and analyzing performance at this level can one begin to compare the models. Of course, what remains is not a study of the models, but a skeletal version of the original, without the modeler-supplied life support systems needed to produce successful forecasts.

[5]Just as we are convinced that a new user would have to live with LITMUS or LITMUS II for 3–6 months before really understanding it, we are also convinced that we would have to live with other sophisticated models such as ASSESSOR or Dodson–Muller for several months before becoming familiar with them.

COMMENT

MARKETING SCIENCE
Vol. 3, No. 3, Summer 1984
Printed in U.S.A.

H. J. CLAYCAMP, J. DODSON AND N. DOUGHTY
Purdue University
N W Ayer Incorporated

We must confess we were surprised at how well the square root function used by Mahajan, Muller, and Sharma as a proxy for the NW Ayer Model performed despite the difference in the dependent variable (brand awareness vs. advertising content recall), the difference in the basic advertising measure (GRP's vs. Adjusted Household Impressions) and the exclusion of additional independent variables used in the original NW Ayer New Product Model.

It is interesting to note from Table 6 that the models have similar patterns of over and under estimation, i.e., under estimation of brands A1, A2 and B1; and over estimation of brands A3, B2 and C1. In particular, with the exception of the first observation for Brand B2 the sign of the error produced by the square root transformation is consistent for all observations of a given brand. This suggests to us that a brand specific effect has not been captured by the test equation. We wonder what the results would show if the test included a measure of product positioning for each brand as an independent variable.

We feel that an explanation for the relatively good performance of all the models examined is the result of (1) the insensitivity of brand awareness as a measure of the impact of advertising and (2) the limited range of the data used to fit the models.

The estimated constant for the square root transformation implies $A_0 = 0.39$. The estimated values of A_0 for the other models ranged from 0.45 to 0.50. Such high levels of "yea-saying" are one of the reasons why the builders of the original NW Ayer New Product Model used a more sensitive measure of the impact of advertising exposure on new product introductions, namely "Proven Advertising Recall."

Actually, within the range of the data, a straight line fits just as well as any of the

0732-2399/84/0303/0201$01.25

models. Consider these two linear fits:

	Brand Data	Category Data
Constant	0.49736	0.54412
Coefficient	0.00004	0.00015
r^2 (adj.)	0.956	0.488
Mean Absolute Error	0.01	0.06

We feel that the appropriate conclusion from this study should be revised to read "all the models provide pretty good fits for both the brand and the product category" *within the limited range of the data used.*

COMMENT

MARKETING SCIENCE
Vol. 3, No. 3, Summer 1984
Printed in U.S.A.

JOE A. DODSON, JR.
N W Ayer Incorporated

I am pleased by the performance of the Dodson and Muller model. It is particularly impressive in light of the restrictions placed on the model. The model was developed to explicitly introduce the effect of word-of-mouth into a general model of the diffusion process. Mahajan, Muller, and Sharma have reduced the model to a conveniently estimatable form which excludes the dual effects of advertising and word-of-mouth.[1]

Mahajan, Muller and Sharma have sacrificed much of the value of the more generalized models they evaluate by restricting their comparison to reduced forms which can be estimated by ordinary least squares. This restriction has stripped the Dodson and Muller model of its richer representation of the adoption process. Perhaps the relatively good performance of the stripped-down version of the model may be explained by the diminished importance of word-of-mouth effects for low priced, frequently purchased, branded products. For these types of products, the cost of trial is nominal, thus the important role of word-of-mouth in information search is diminished.

Mahajan, Muller and Sharma describe the Dodson and Muller model as excluding the effects of promotion. Marketing instruments, such as sampling, couponing, and distribution stimuli, though not explicitly represented in the model, are represented by the μ coefficient in the original Dodson and Muller article. We did not specify a functional relationship between it and the firm's marketing instruments. Mahajan, Muller, and Sharma have assumed that the only effect is due to advertising as measured by GRP.

Finally, the fact that the comparisons made by Mahajan, Muller, and Sharma were forced by the data available is an indication of the lack of sophistication with which most companies monitor their investments in new products. Where models are used, attention should shift from simple prediction to strategy development. Good models force consideration of all aspects of the marketing problem at hand and should provide help in answering questions of "why" not just "how much."

[1] An unabridged version of the original Dodson and Muller article (available from the TIMS office) does describe and illustrate a procedure for obtaining empirical estimates of the complete model's parameters, including the word-of-mouth effect.

0732-2399/84/0303/0202$01.25

COMMENT

MARKETING SCIENCE
Vol. 3, No. 3, Summer 1984
Printed in U.S.A.

CLARIFICATION OF THE TRACKER METHODOLOGY AND LIMITATIONS

JOHN L. GOLANTY
Golanty & Associates, Inc.

The following comments are designed to clarify several points about TRACKER that may not be obvious from the JMR (5/78) article, yet are relevant to the preceding discussion.

(1) Although sampling was not treated explicitly in our description of TRACKER, it certainly has a major effect on awareness and is a factor we examine very closely. The awareness level derived for a brand using the equations in our article is for *nonsampled category users*. Respondents who remember having received a sample can be added to the nonsampled group to obtain a total awareness figure, but for analytical purposes we always evaluate sampled respondents separately.

(2) The awareness model utilized in TRACKER was designed to capture the effects of television GRPS on total brand awareness for brands that are *heavily* advertised. High levels of television advertising tend to mask the importance of other variables because the medium is so pervasive and influential. For new brands supported by low levels of TV weight (or no TV), factors such as distribution, in-store promotional activity, and word-of-mouth communication assume far greater importance in the generation of awareness. The problem for the researcher is how to accurately measure these variables at an affordable cost—especially since unadvertised brands tend to have lower overall budgets (including dollars allocated for research).

(3) In order for category data to provide good pooled results, we find we usually have to define a category quite narrowly. For example, light beer and superpremium beer generate very different results. Awareness for each of these subgroups cannot be estimated well by grouping together all brands of beer. The extent to which brands can be grouped together is a function of the homogeneity of the target consumer, the ways in which the new products are likely to be used, and the similarity of the marketing plans.

COMMENT

MARKETING SCIENCE
Vol. 3, No. 3, Summer 1984
Printed in U.S.A.

ISSUES IN COMPARING THE AWARENESS COMPONENT OF NEW PRODUCT MODELS

LEWIS G. PRINGLE, R. DALE WILSON AND EDWARD I. BRODY
BBDO, Inc.
Michigan State University
BBDO, Inc.

Mahajan, Muller, and Sharma (MMS) have provided the marketing community with the first published attempt at empirically comparing one aspect of alternative new product models. Their comparison of several models using the same data is a positive step and we believe that the authors are to be congratulated for their efforts. At the

0732-2399/84/0303/0203$01.25

same time, several issues remain to be addressed before firm conclusions can be drawn regarding the efficacy of each tested model.

1. First, the assumptions used in their comparison have the effect of transforming dissimilar models into sets of equations quite similar in structure for the purposes of the empirical test. These assumptions eliminate rather substantial inherent differences among the models and, as a result, only a partial comparison was in fact achieved. For example, advertising was considered in the authors' restatement of the NEWS model to be the only source of brand awareness while promotional data (i.e., sampling and couponing activity) were excluded from the analysis. This omission is relevant since total promotional expenditures often outweigh total advertising expenditures during the introductory phase of a typical packaged good. The data actually submitted included a media coupon with 33% reach and a mail coupon with 30% reach. As a result of this assumption (and others) imposed for the purposes of generating comparable results, it is not surprising that the models appear to perform similarly in the empirical comparison. They are virtually forced to do so.

2. The treatment of initial awareness (A_0) raises another important issue. By estimating A_0 from the data, MMS found values ranging from 0.45 to 0.50 for all models tested. For NEWS, the estimated values were both 0.45. As the authors note, such high values for initial awareness are not consistent with "new products." Their explanation of "A_0" as capturing effects prior to the first observation creates problems in that the conceptual distinction (in both TRACKER and NEWS) between awareness prior to advertising, thus not caused by media exposure, and awareness caused by media exposure becomes obliterated. It is also not indicated whether the first observation is the first advertising period for the TRACKER data. It seems to us that MMS *have* confirmed that A_0 should *not* be treated as a fitted parameter, even if this means a slightly less accurate fit to the data. A_0 *is* observable and it *should* be observed. If not, why should one not continue the sequence with estimates of $A_1, A_2, \ldots$ as parameters? In short, it seems to us that specification accuracy has been sacrificed in favor of goodness-of-fit.

3. The question of using pooled product class data for the purposes of parameter estimation highlights a major philosophical difference between NEWS and TRACKER. The two modeling approaches have been fully contrasted elsewhere (Pringle, Wilson, and Brody 1982, p. 7; Wilson and Pringle 1982, p. 308). Consider a physician examining a patient. Wouldn't tests on the current patient be relied on more heavily than the pooled tests of his six previous patients? We believe that, like a prudent physician, the new product analyst must examine the specific case. We find that prior product class information is useful as a judgmental reference, but not as a direct input to the NEWS model. We also note that the TRACKER model does *not* use pooled data for its repeat submodel.

4. Concerning the role of word-of-mouth communication, our experience has shown that, for the types of new consumer packaged goods for which NEWS is applied, word-of-mouth is far less important than other sources of market information. Word-of-mouth is no doubt much more important for durables, other high ticket products having a lengthy purchase cycle, products having a low advertising/promotion budget, and for products having a high degree of psychological risk. Since these conditions are not met by most packaged goods, we believe that word-of-mouth plays a generally insignificant role. As pointed out by Blattberg and Golanty (1978, p. 194), the cost of measuring word-of-mouth communication would be prohibitive relative to its value in a model designed for most packaged goods.

5. The treatment of *forgetting* by MMS requires that additional work be undertaken in subsequent investigations. Forgetting, in new product modeling, is a function of the length of the brand purchase cycle as well as the media plan used during the

introductory period. For example, a new brand having a short purchase cycle (e.g., two weeks) and a steady accumulation of GRP's would be subject to minimum awareness decay in each period. On the other hand, if the purchase cycle is longer (e.g., three months) and the media plan calls for a *flighted* schedule, substantial decay in awareness would be observed. The applications reported by MMS indicate that their results do not seem to be sensitive to changes in the forgetting parameter. However, applications using other data sets do indeed indicate a more significant contribution of this phenomenon.

Despite the above reservations, we believe that MMS have provided a commendable first step. We look forward to further studies of this type using a more faithful model specification and richer data bases. This approach should also be extended to include the purchase variables: trial, repeat and sales.

References

Blattberg, R. and J. Golanty (1978), "Tracker: An Early Test Market Forecasting and Diagnostic Model for New Product Planning," *Journal of Marketing Research*, 15 (May), 192–202.

Pringle, L. G., R. D. Wilson and E. I. Brody (1982), "NEWS: A Decision Oriented Model for New Product Analysis and Forecasting," *Marketing Science*, 1 (Winter), 1–30.

Wilson, R. Dale and Lewis G. Pringle (1982), "Modeling New-Product Introductions: A Comparison of NEWS, SPRINTER and TRACKER," in *Analytic Approaches to Product and Marketing Planning: The Second Conference*, Rajendra K. Srivastava and Allan D. Shocker (Eds.), Cambridge, Mass.: Marketing Science Institute, 297–311.

REJOINDER

MARKETING SCIENCE
Vol. 3, No. 3, Summer 1984
Printed in U.S.A.

REFLECTIONS ON AWARENESS FORECASTING MODELS OF NEW PRODUCT INTRODUCTION

VIJAY MAHAJAN, EITAN MULLER AND SUBHASH SHARMA
Southern Methodist University
The Hebrew University of Jerusalem
University of South Carolina

We are flattered to have the developers of the various models comment on our paper. Needless to say, we are indebted to them for their valuable comments. Some of the comments, however, merit further discussion and our reaction to these comments is provided below.

For example, how can we argue with Blackburn and Clancy when they say that they like our paper because "the results are favorable to our model, LITMUS" and "the paper convinces us that our recently introduced LITMUS II model would perform even better?" On the other hand, they say that "this evaluation appears to be based on data neither LITMUS nor any of the other models was designed to forecast" or "when data on the performance of a plan are available for the first six months, it takes neither a genius nor a complex model to forecast year-end with reasonable accuracy." Even though the scope of our paper is modest, the problem is real, the data are real, the models are real, and the empirical results are real. Given the lack of details on the analytical developments and the sources of data for LITMUS (as acknowledged by Blackburn and Clancy), we thought we did a good job of deciphering their model. Also, wherever appropriate, we gave them the benefit of the doubt by mentioning that LITMUS acknowledges the inclusion of this factor or that factor or "we understand

0732-2399/84/0303/0205$01.25

from its developers that since the model actually evolved from NEWS, its data sources and estimation procedures are similar to NEWS."

We also appreciate the further clarification and limitations of the TRACKER model provided by Golanty. Similarly, regarding the Dodson/Muller model, we are indebted to Dodson for reminding us that "perhaps the relatively good performance of the stripped-down version of the model may be explained by the diminished importance of word-of-mouth effects for low priced, frequently purchased, branded products."

We don't quite understand why Claycamp, Dodson and Doughty are "surprised" at how well the square root transformation used in the AYER model captures the dynamics of the relationship between advertising and awareness "despite the difference in the dependent variable (brand awareness vs. advertising content recall), the difference in the basic advertising measure (GRP's vs. Adjusted Household Impressions) and the exclusion of additional independent variables used in the original Ayer New Product Model." Concerning the dependent variable, it is certainly true that brand awareness is *only* one aspect of "content recall." However, since they define the dependent variable, AR, to be the percentage of consumers who are able to "accurately" recall advertising claims, it is very likely that there is a strong correlation between the two measures. With respect to the independent variable measuring advertising, we understand from them that Adjusted Household Impressions (AHI) is *not* GRP. Instead, in order to develop AHI, they use a "Media Equalizer Process" that adjusts GRP based on the specific vehicles used. We hypothesize that there is a strong correlation between AHI and GRP. However, even if GRP does not "efficiently" represent AHI, should the nature of the relationship (diminishing returns) be different? The AYER model certainly uses the square root transformation to represent this effect. We agree with their concern about the exclusion of other variables such as product positioning in the model. However, given the nature of the data we had and some of the empirical results (e.g., Tables 2 and 8), we really wonder how much additional variance in the data would have been explained by the inclusion of these additional variables. We are very surprised, however, that Claycamp, Dodson and Doughty feel that a straight line may be appropriate for these data. That is, $A_t = \alpha + \beta(\sum \mathrm{GRP}_i)$. Consider, for example, the results for the category data reported in their comments, i.e., $\alpha = 0.544118$ and $\beta = 0.0001508$. Since the linear model, like the AYER model, does not impose a ceiling on the maximum level of awareness, for the last observation of the seventh brand (brand C_2 in Table 8), the linear model predicts awareness of 109% (?) for cumulative GRPs = 3600. On the other hand, the square root model predicts the actual awareness fairly well. We must, therefore, disagree with their conclusions.

The comments provided by Pringle, Wilson and Brody highlight the major differences between the model development philosophies of NEWS and TRACKER. The limited empirical evidence in our paper indicates that estimation of initial awareness can generate relatively better fits and the data can be pooled to generate brand awareness. However, our results are based on only two data sets and exclude other awareness generating stimuli. Hence, any generalization on our part on the merits of the basic philosophies of these two models would be premature. Clearly, there is a need to determine when it is appropriate to use one philosophy versus the other.

Finally, we echo Pringle, Wilson and Brody in their call for more empirical studies using "faithful" model specifications and richer data bases, including other purchase variables such as trial, repeat and sales. Given the preponderance of new product forecasting models, our discipline can definitely benefit by comparative studies examining the analytical similarities and performance differences among the various models. We hope that our study will be viewed as a first step towards the achievement of such a goal.