The Marriage Of Graphic Design & Research – Experimentally Designed Packages Offer New Vistas & Opportunities

Richard Bernstein, Ph.D.
Howard Moskowitz, Ph.D.

Moskowitz Jacobs Inc.
1025 Westchester Avenue
White Plains, New York
Phone: 914-421-7400
Fax: 914-428-8364
e-mail Rbernste@MJI-Designlab.com

For many fast moving consumer goods the package is the first experience a consumer has with the product. As part of MJI's services, we apply advanced conjoint (Green & Srinivasan, 1978) methods to help our clients determine best packaging for their products. In some of our work, we help clients develop the most visually appealing package (container) for their product. In other studies, we help our clients develop the most appealing graphics to decorate the their packages. The work we do owes much to a psychophysical heritage (e.g. Moskowitz, 1981) – systematic stimulus manipulation is coupled with a simple respondent-rating task. Models relating visual appeal to packaging options are developed.

We will describe case studies, which demonstrate two applications of our approach:

1) Package-design studies in which the goal is to maximizing the appeal of the physical container. In these studies, basic physical parameters such as package shape, dimensions, the surface material, opening and closing mechanisms are varied. Respondents are exposed to photo-realistic renderings of the packages. From these studies, our clients understand the impact of each packaging option on consumer appeal.
2) Graphics-design studies in which the goal is to maximizing the appeal of package graphics. In these studies, graphical design elements such as Logo's, backgrounds etc are systematically varied. Target consumers are exposed to photo-realistic drawings of the packages embellished with the graphical elements.

In both applications, we work with our clients to develop parameters that both vary along physically meaningful dimensions and yield viable experimental designs.

Once the packaging has been parameterized, we are in a position to develop our stimulus array (cf., Moskowitz, 1981). We then ask respondents to rate an experimentally designed set of packages on one or two key attributes such as interest, liking, purchase intent, etc. Because the test array conforms to an experimental design, dummy variable multiple regression can be used to develop analytical models estimating each packaging option's contribution to package appeal. Armed with these models, our clients have the tools necessary to develop packaging optimized for their marketing goals.

Case Study #1 Basic Package Parameters

Background

We illustrate our package-design work with a study commissioned by a multi-national manufacture reviewing their packaging strategy. The manufacturer was seeking to understand the packaging features that stimulate consumer interest and motivate purchase. The company wanted to know the impact and relevance of a variety of packaging options in order to develop the best packaging for their different user groups. Of particular interest to our client was examining the feasibility of developing a multi-national strategy. The study was executed in 4 separate European markets. The respondent sample was constructed so that separate models could be developed all important user sub-groups. Specifications for the sub-group were based on client's marketing objectives.

Stimuli

The manufacturer's design-team identified all the viable packaging options. Varied were the package shape, opening and closing mechanisms, surface material as well as the presence/absence of several value-added features. There were 4 types of package shape—1 type that was rigid and 3 types had flexible construction. Since package shape and closing mechanism were highly correlated, these characteristics were combined into a single parameter. In total, there were 42 packaging options organized into 5 categories of related characteristics: A fractional-design (e.g. Box et. al., 1978) was used to specify 186 test packages. Each of the test packages was realizable (i.e. could be manufactured). A design-firm generated photo-realistic depictions of each of the test packages on computer readable media. Package visual consisted of a main picture of the package on the left and supplementary inserts of the key features and descriptive text of the features on the right. Text was translated into the language appropriate for the test market. (See Figure 1.)

[Insert Figure 1 about here]

Additional test packages were developed to assess the impact of specific packaging options not suitable for inclusion in the experimental. Since these "overlays" addressed issues that are client-specific and proprietary, details will not be discussed in this paper.

Method

<u>Recruit</u>
In accord with local practices, respondents were pre-recruited via telephone or "intercepted". All study material was presented in the local language. After completing a screening questionnaire, qualified respondents were invited to participate in the study at a central location. Qualifications included demographic and product usage criteria. Each session included 1-10 respondents.

Session Sequence

Orientation

Respondents were seated at individual computer workstation. At the beginning of the session, respondents participated in a short orientation session. Moderators were provided with the orientation script in which the various package options were described and then the respondents handled each of 15 examples. This allowed the respondents to "experience" the various packaging options explored. This orientation was intended to help respondents appreciate the packages that were depicted in the main study.

The moderator then took respondents through an exercise to familiarize them with the use of the computer keyboard, the study questions and the continuous 9-point rating scale. They were encouraged to use the entire 9-point rating scale.

Package Positioning

In order to provide a context for the rating exercise, respondents were provided with a positioning statement that read:
"You will be looking at a variety of packaging alternatives for the product category. Some of the features you see may be familiar and some are new ideas. In any case the package is colored in a neutral color for this test. Do not use the color as a signal for the product quality.

When the computer shows you a package, look at it carefully and read all of the written descriptions. Assume the product you use most often at home is in the package.

Please take your time and read each concept (screen) thoroughly. Now, tell us how you feel about the package using the questions shown on the bottom of the computer screen. Enter your rating based on the following questions. The entire concept should be rated as a whole."

Phrases in Italics were replaced with appropriate category-specific references.

Package-Rating Exercise

Trials were self-paced. Package visuals were presented on a computer screen and respondents rated how well they liked the package and also they rated each package on an additional attribute. Short rest periods were imposed after sets of 25 packages were rated. Figures 1 and 2 present sample screens created for illustration and were not part of the study. Note that all the packages and backgrounds were uniformly colored. Actual insert describing features were in the appropriate language and were supplemented with an icon representing the feature. Both the feature inserts and the orientation excise was intended to enrich the representation of test packages. Each respondent rated 90 package concepts. They rated 78 of 186 test packages, and all 12 "overlays". Presentation order randomized and care was taken to balance package options within and between respondents.

[Insert Figures 2 about here]

Classification questionnaire

After completion of the rating task, a short classification and usage questionnaire was administered. Responses to this questionnaire are used to develop additional respondent classification schemes. In total a test session lasted about 45 minutes.

Results

Following market research convention, data is organized so that we can estimate, percentage of respondents giving top-3 box ratings (i.e., ratings of 7,8 or 9) for each of the tested packages. Given the experimental design matrix, we compute a dummy variable regression of the form:

$$Y = b_{(0)} + b_{(c1,1)}X_{(c1,i)} + \ldots + b_{(cl,1)}X_{(cl,j)}.$$

Where, Y is the package liking score (or other rating attribute) and each X represents the presence/absence of the package options. The indexing variable indicates the option's category membership. Since the design matrix was constrained to contain only realizable packages, the regression coefficients are estimates part-worth contribution of each packaging component. We build separate models for each analytical group.
Every package has a shape/closing, opening mechanism and a material. Because an option from these categories is always present, we identified one option as a reference option. Coefficients for these categories represent the contribution relative to the reference. For convenience, the current (in-market) packages were assigned as the references. Thus computed coefficients indicated the packaging options impact on liking relative to current in-market packages. Positive coefficients indicate that the option increases liking, negative coefficients indicate that the option yields less liking than the current option. Coefficients near 0 indicate that option is at parity with the in-market package. Since our models are linear, the reference option can be changed without impacting the liking-hierarchy. Value-added features, in contrast, yield coefficients that represent absolute incremental contribution to liking (presence vs. absence) .

To examine how well the regression model describes consumer liking, package-liking scores are estimated from model parameters and compared to measured liking scores. Figure 1 presents package-liking for the total sample. Clearly the fit is good. Indeed, individual countries yielded r^2 ranging from 0.86 to 0.93. Thus our regression models provide an accurate descriptions of how package-liking varies with packaging parameters.

[Insert Figure 3 about here]

Table 1 presents the models for the total sample and for each of the 4 European markets. Indicated are the regression coefficients for each package based on models calculated for each market Displayed are $b_{o,}$ the model constant, which absorbs, consumers generalized liking and their use willingness to use the high-end of the rating scale. Option-coefficients indicate each options contribution to liking. Options with positive coefficients indicate that packages with that option yield greater liking than the corresponding reference options. Options with negative coefficients are liked less than the reference option. Options with coefficients near zero yield liking at parity with the reference option. Options are ordered by their contribution to liking of the total sample model.

[Insert table 1 about here]

It is striking that the 4 countries yield similar models. Option preferences have similar rank ordering. Options that are preferred in one country tend to be preferred in all countries. Consider first, the findings on package shape/closure. Several Type A packages are liked the best. This finding is particularly intriguing since the different regions have different package histories and different in-market packages.

Consider next the opening method. In every region the same opening method (method A) is preferred. Interestingly, the impact of opening method on liking is about the same for all user groups. In contrast, consumers do not show any strong or consistent preference for material. Package appeal is impacted strongly by shape of the package and how it is opened and closed. In contrast, appeal is only minimally impacted by material.

Finally, consider the two value-added features. Both are intended to increase the usability of the package. Both contribute to consumer liking and so seem to be valued by the consumer. As part of usage questionnaire, we asked consumer to indicate how much they were willing to pay for each of these value-added features. We used that series of questions to classify a respondent as being willing or unwilling to pay for the feature. We then built liking models for consumers *unwilling* to pay extra for any features and also for consumers *willing* to pay extra for both features. Table 2 presents coefficients for the value-added features. Consumers unwilling to pay extra for features like the features less than consumers who are willing to pay for the features. This finding, although not surprising, does illustrate an internal consistency in the data structure.

[Insert table 3 about here]

US Replication

Quite recently, we were commissioned by the US affiliate to execute a similar study in the United States. Package options under consideration were similar (but not identical) to those under consideration in our European study.

Study Design

Once again we systematically varied basic package parameters. Because package shape determines realizable method of closure, we again combined these dimensions into a single design variable. Method of opening and surface material options was similar to the European study. In total, there were 126 test-packages representing a full-factorial design of all design-variables. Each respondent participated in a scripted orientation and rated 75 test-packages as well as several "overlays". As in the European study, a short classification questionnaire was administered after the package rating exercise.

Results

Figure 4 presents a comparison of obtain and modeled package-liking scores. Clearly consumers discriminate between packages. Best liked packages yield scores about 70 whereas the least liked packages yield scores near zero. Modeled scores capture consumer preferences.

[Insert Figure 4 about here]

The model computed for the total sample is presented in Table 4. Two findings are compelling. Consumers in the US have a preference hierarchy similar to their European counterparts. Packages have near-universal preference hierarchies. Given the type of product, we suspect that the preferences are driven by feature-functionality.

There is one difference between European and US findings. In the US there is a clear preference for Type-C material. We note that Type-A packages are virtually always made of Type-C material in the US, whereas, in Europe, Type-A packages are made from a variety of materials. It is intriguing that this US-European difference may reflect difference in consumer experience.

[Insert Table 4 about here]

Case Study #2 Package Graphic

Background

The second application is illustrated by a work commissioned by a domestic manufacturer considering re-launching one of their product lines. A key goal was to identify the most appealing package graphics. Graphical components were varied in order to identify the packaging graphics with the strongest consumer appeal. Under consideration were three product brand names, two background designs, two package colors and three product descriptions (quips).

Experimental Design

A full-factorial design was used to define the test-array. Each representation consisted of a photo-realistic rendering of the front of the package and an insert showing an enlargement of a key section of the package. Care was taken to ensure that all text was legible. Each respondent rated all 36 test packages as well as small set of "overlays". Figures 5 and 6 illustrate the type of graphical variations explored. These illustrations were created for this paper and were not used in the study.

[Insert Figure 5 about here]
[Insert Figure 6 about here]

Results

Figure 7 presents a comparison of obtain and modeled purchase interest scores. We that consumers show clear preferences and that the model successfully captures consumer ratings. Table 5 presents the model for the total sample. Best graphic options are readily identified by inspection of component coefficients.

Summary

We help our clients maximize the appeal of their packaging. We developed experimentally designed test arrays that vary along physical dimensions. Arrays are constrained to contain only realizable packages. Targeted consumers are exposed to the test-array and are asked to provide ratings of package appeal. Our analysis decomposes ratings and assigns contribution scores for each packaging option. The sample is selected so that we can develop separate models of appeal for each important marketing group. In effect, we use an experimentally defined array to determine the impact of each packaging option for all important marketing target groups. This allows our clients to readily determine best packaging option for a variety of marketing scenarios.

References

Box, G.E.P., Hunter, J. & Hunter, S. 1978. Statistics for Experimenters, John Wiley, New York.

Green, P.E., & Srinivasan, S. 1978 Conjoint analysis in consumer research: Issues and outlook. Journal of Consumer Research, 5, 103-124

Moskowitz, H.R. 1981. Psychophysical approaches to package design and evaluation. In Handbook of Package Design Research. W. Stern (Ed.), p.505-534. John Wiley, New York

Table 1 Additive Model for Package Liking. Entries are part-worth contribution for packaging components.

		European Country			
	Total Sample	**1**	**2**	**3**	**4**
BASE SIZE	***811***	***204***	***207***	***200***	***200***
b_o	**25**	**22**	**33**	**16**	**41**
PACKAGING SHAPE					
TYPE A-1	21	26	17	29	5
TYPE A-2	19	27	12	27	3
TYPE A-3	18	21	10	31	2
TYPE A-4	16	17	16	27	-1
TYPE A-5	16	23	9	26	0
TYPE A-6	14	16	10	26	-2
TYPE A-7	12	15	6	24	-4
TYPE A-8	9	9	5	20	-7
TYPE B-1	6	10	1	19	-11
TYPE C-1[REFERENCE country 4]	3	1	0	2	0
TYPE B-2	2	3	-1	12	-14
TYPE C-2 [REFERENCE country 1-3]	0	0	0	0	-7
TYPE B-3	0	-7	-2	10	-9
TYPE C-3	-3	-5	-7	2	-7
TYPE C-4	-3	-8	-3	0	-9

TYPE C-5	-4	4	-11	6	-21
TYPE C-6	-5	1	-9	1	-20
TYPE D-1	-5	-9	-8	3	-13
TYPE B-4	-5	-5	-11	3	-14
TYPE D-2	-6	-5	-9	1	-18
TYPE C-7	-7	-5	-10	0	-22
TYPE C-8	-8	-7	-12	-1	-17
TYPE C-9	-10	-14	-15	-5	-13
TYPE C-10	-11	-12	-15	-4	-17
TYPE C-11	-11	-11	-18	-1	-21
TYPE C-12	-12	-14	-17	-4	-21
TYPE D-3	-12	-13	-17	-4	-23
TYPE C-13	-13	-15	-18	-6	-19
TYPE C-14	-13	-16	-15	-8	-20
TYPE D-4	-16	-6	-20	-14	-31
TYPE C-15	-22	-16	-28	-19	-30

Table 1 (Continued)

		European Country			
	Total Sample	**1**	**2**	**3**	**4**
Opening Method					
Method A [REFERENCE - country 4]	8	8	7	9	0
Method B [REFERENCE -country 1-3]	0	0	0	0	-8
MATERIAL	0	0	0	0	4
TYPE A [REFERENCE]	0	1	2	-3	3
TYPE B	-1	1	-1	0	0
TYPE C	-1	1	-1	-2	1
TYPE D	-1	1	-1	-1	0
TYPE E					
ADDED VALUE FEATURE #1					
Presence	7	5	6	11	4
Absence [REFERENCE - all countries]	0	0	0	0	0
ADDED VALUE FEATURE #2					
Presence A	7	4	6	14	6
Absence [REFERENCE - all countries]	0	0	0	0	0

Table 2 Additive Model for Package Liking. Entries are part-worth contribution for packaging components. Separate models were built for consumers indicating an unwillingness to pay extra for any value added features and for consumers indicating a willingness to pay extra for both features.

	Total Sample	Willing to Pay Extra for All Features	Not Willing to Pay Extra for Any Feature
BASE SIZE	**811**	**339**	**157**
b_o	**25**	**25**	**28**
ADDED VALUE FEATURE #1			
TYPE A	7	8	3
TYPE B [REFERENCE POINT]	0	0	0
ADDED VALUE FEATURE #2			
TYPE A	7	8	3
TYPE B [REFERENCE POINT]	0	0	0

Table 3 Additive for US Replication. Preferences are quite similar to those found in Europe. Consumers react similarly to basic package parameters.

	Total Sample
	Liking
BASE SIZE	**387**
B_0	**-1**
Package Shape/Closure	
TYPE A -1	55
TYPE A -2	50
TYPE A -3	49
TYPE A -4	49
TYPE A -5	47
TYPE A -6	47
TYPE C -1	27
TYPE C -2	26
TYPE C -3	24
TYPE C -4	21
TYPE C -5	20
TYPE C -6	20
TYPE C -7	19

TYPE C -8	18
TYPE C -9	16
TYPE C -10	14
TYPE D -1	10
TYPE C -11	9
TYPE C -12	8
TYPE C -13	7
TYPE D -2	6
TYPE C -14	6
TYPE C –15	3
TYPE D -3 [REFERNCE]	0

Table 3 Continued

	Liking
OPENING	
TYPE A	9
TYPE B [REFERNCE]	0
MATERIAL	
TYPE C	7
TYPE A	4
TYPE B [REFERNCE]	0

Table 4. Additive model for Graphical Elements.

	Total Sample
BASE SIZE	**103**
B_o	**54**
Designs	
TYPE A	10
TYPE B [REFERENCE POINT]	0
Names	
Name 1	7
Name 2	4
Name 3 [REFERENCE POINT]	0
Colors	
TYPE A	5
TYPE B [REFERENCE POINT]	0
Product description (Quip)	
TYPE A	3
TYPE B	0
TYPE C [REFERENCE POINT]	0

Figure 1. An example of screen representation of a package. On the left, a photo-realistic rendering of the package. On the right, inserts highlighting key features. On the bottom, a prompt for the rating question.

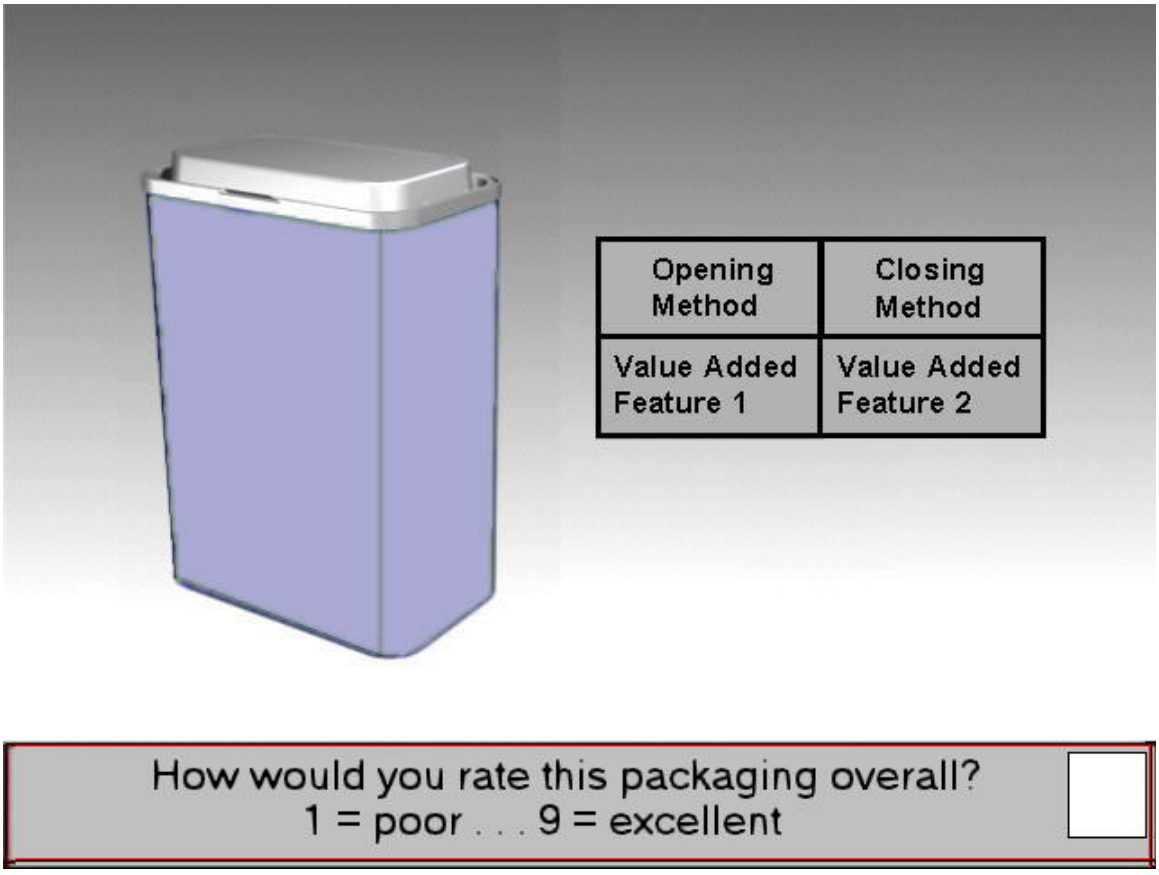

Figure 2. An other example of screen representation of a package. On the left, a photo-realistic rendering of the package. On the right, inserts highlighting key features. On the bottom, a prompt for the rating question.

Opening Method	Closing Method
Value Added Feature 1	

How would you rate this packaging overall?
1 = poor . . . 9 = excellent

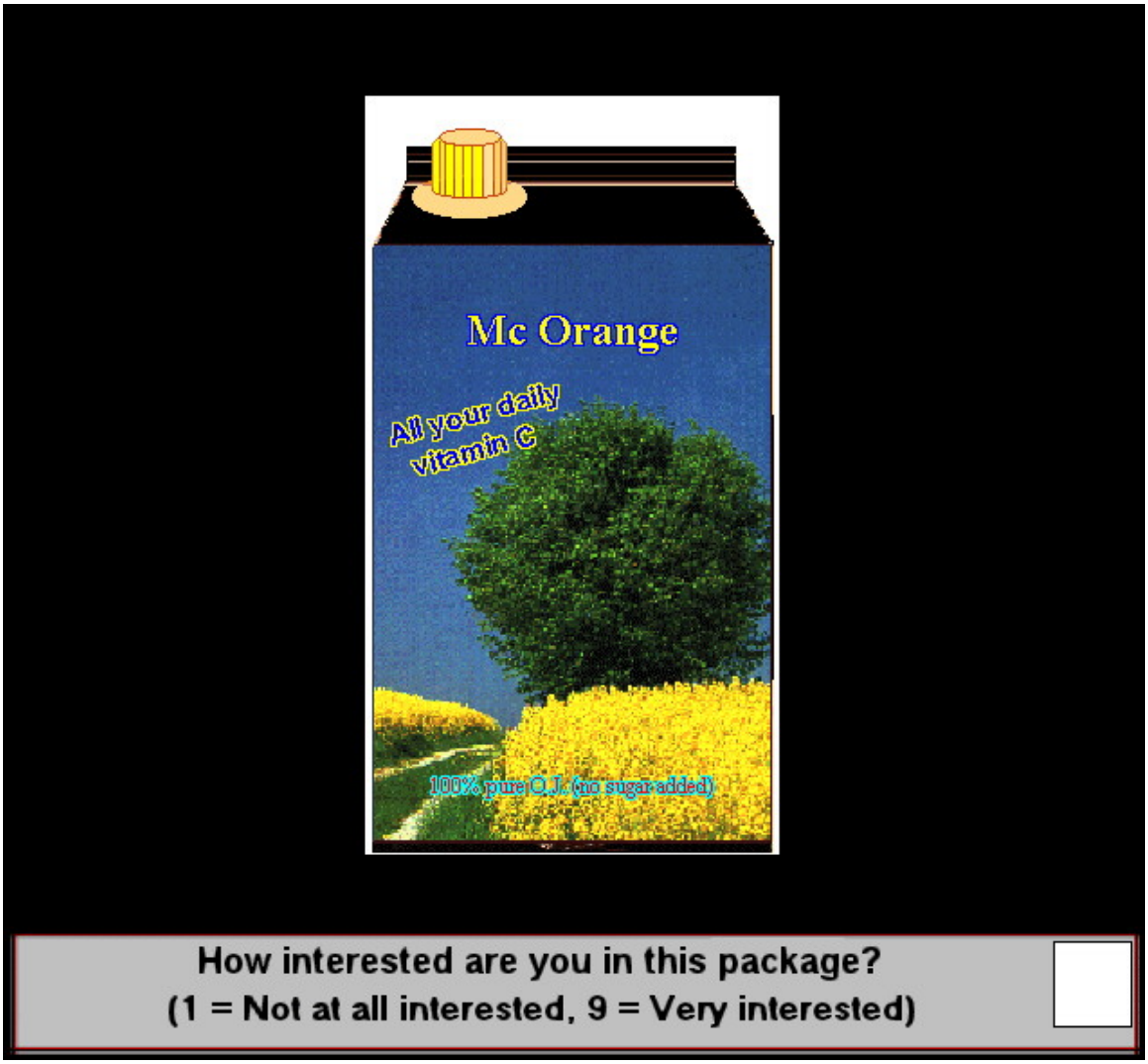

Figure 5. Illustration of Screen representation of package graphics.

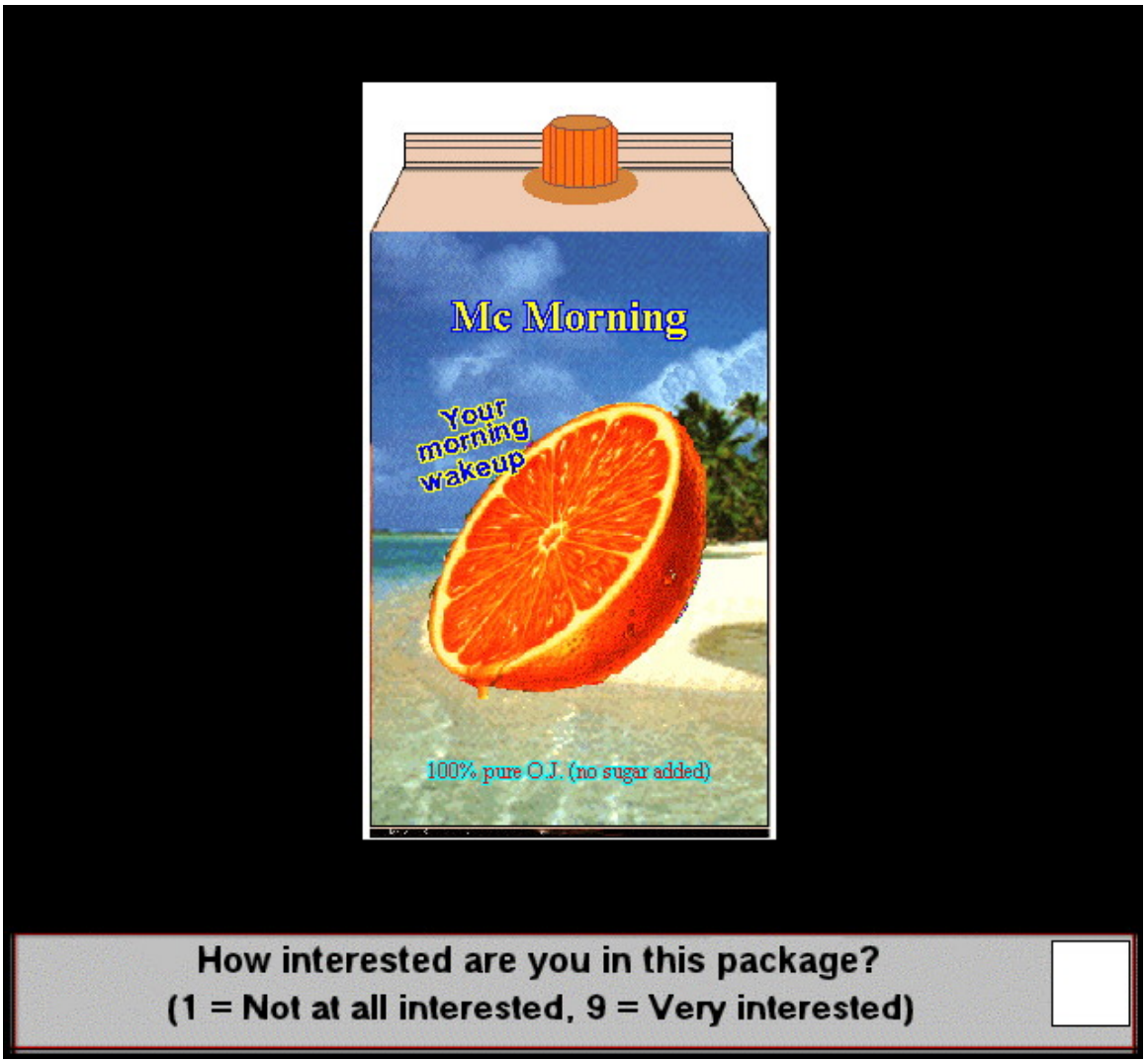

Figure 6. Illustration of screen representation of package graphics. Although the package graphics are systematically varied, each screen displayed a realistic package.

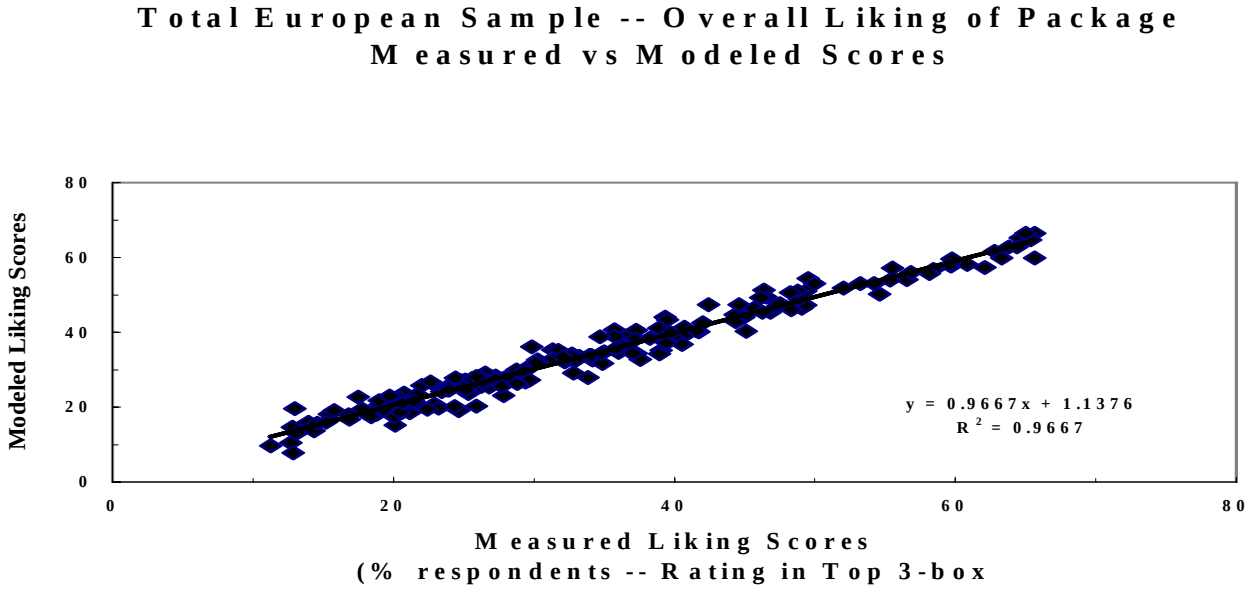

Figure 3 Measured vs. Modeled Liking Scores. The computed regression model accurately describes measured package liking.

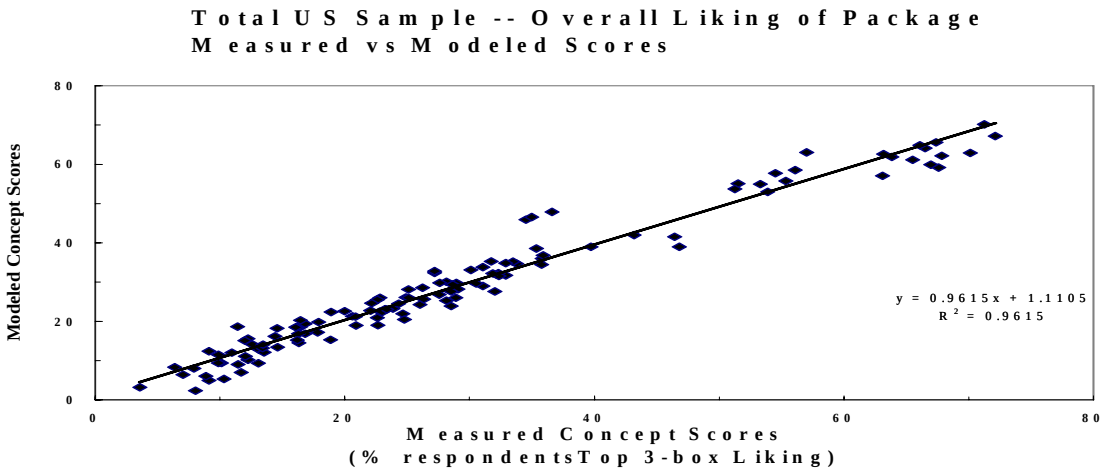

Figure 4 Measured vs. Modeled Liking Scores.

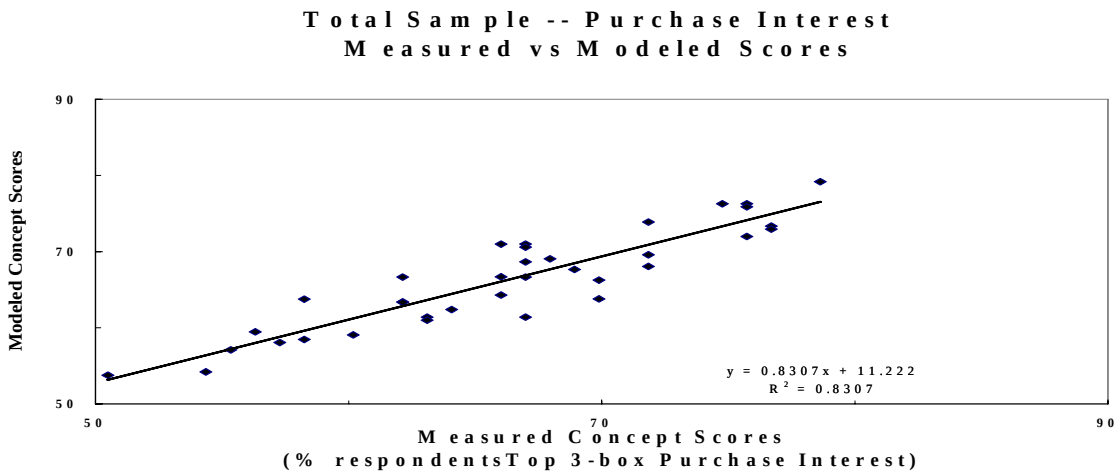

Figure 7 Measured vs. Modeled Purchase Interest Scores.