



.... RESEARCH NOTES AND COMMUNICATIONS

CONTENTS

The Trouble With Concept Testing	<i>Russell I. Haley and Ronald Gatty</i>	230
Correcting Nonresponse Bias In Mail Questionnaires	<i>Peter Ognibene</i>	233
Semantic Properties of Selected Evaluative Adjectives: Other Evidence	<i>Robert A. Mittelstaedt</i>	236
Semantic Differential and Nonmetric Multidimensional Scaling Descriptions of Brand Images	<i>Joseph N. Fry and John D. Claxton</i>	238
Is There a Generalized Price-Quality Relationship?	<i>David M. Gardner</i>	241
Personality and Innovation Proneness	<i>Jacob Jacoby</i>	244
"Psychophysics of Prices": A Reappraisal	<i>Kent B. Monroe</i>	248
Comments on "Psychophysics of Prices"	<i>André Gabor, Clive W. J. Granger, and Anthony P. Sowter</i>	251
"Psychophysics of Prices": A Reaffirmation	<i>Joseph M. Kamen and Robert J. Toman</i>	252
Overlap of Opinion Leadership Across Consumer Product Categories	<i>Seymour Sudman</i>	258
Overlap of Opinion Leadership: A Reply	<i>John O. Summers and Charles W. King</i>	259

The Trouble With Concept Testing

RUSSELL I. HALEY and RONALD GATTY*

THE NATURE OF THE PROBLEM

The terms "concept" and "concept testing" have been used as loosely as any words in the vocabulary

* Russell I. Haley is Executive Vice President of the Grudin/Appel/Haley Research Corporation. Ronald Gatty is Associate Professor in the Baruch College, City University of New York.

of marketing and advertising. The trouble with concept testing really begins with a lack of definition as to what a concept is.

We consider a concept as an underlying idea of marketing or advertising strategy in planning for new products, repositioning established brands, and developing advertising campaigns. Most importantly, for clarity, the underlying idea must be distinguished from the

specific way in which it is or could be carried out, which is the execution of the concept. The execution may be a physical prototype of a product concept, it may involve a prototype of the package as well as the advertising copy, or it may be restricted to copy alone.

This distinction between concept and execution is of paramount importance. The concept may be good and the execution bad, or vice versa. But the difficulty in concept testing is that tests can evaluate the concept only as interpreted through the specific executions with which it interacts. With the usual tests made, it becomes impossible to determine if the attempted evaluation of the concept is mainly a function of the way in which it is executed, or if it is an evaluation of the concept itself. In advertising research, much so-called concept testing could just as well be considered as copy testing, since there is no way of separating the effects of the concept and the copy.

We feel that the most suitable way to evaluate a concept is through formal design of an experiment that can evaluate the concept itself and a selection of possible executions separately. We can only evaluate a concept through selected executions, as demonstrated through an experimental study designed and conducted under the direction of the senior author.

AN EXPERIMENTAL EVALUATION OF CONCEPTS AND EXECUTIONS

Four hundred housewives in ten cities were identified as prospects for the new product in question. Each was asked to rate product concepts on a six-point hedonic rating scale from "extremely desirable" to "not at all desirable." Each was shown 24 copy executions (8 different concepts by 3 different copywriters). The executions were written (no graphics) and were prepared on

Table 1
MEAN RATINGS OF THREE COPYWRITERS'
EXECUTION OF EIGHT CONCEPTS

Copy-writers	Concepts								Mean	Range
	1	2	3	4	5	6	7	8		
Six-point hedonic scale, weighted 6, 5, 4, 3, 2, 1										
1.	4.42	4.31	5.00	4.43	4.50	4.47	4.57	4.27	4.49	.73
2.	4.44	4.35	4.99	5.03	4.89	4.96	3.87	4.19	4.59	1.16
3.	3.78	4.22	4.96	4.98	4.68	4.14	4.65	4.25	4.45	1.20
Mean	4.21	4.29	4.98	4.81	4.69	4.52	4.36	4.24	4.51	—
Range	.64	.13	.04	.60	.39	.82	.78	.08	—	—
Six-point hedonic scale, weighted 8, 4, 2, 1, 1, 1										
1.	3.88	3.70	4.68	3.85	3.96	3.86	4.00	3.54	3.93	1.14
2.	3.89	3.78	4.67	4.72	4.54	4.74	3.26	3.51	4.14	1.48
3.	3.04	3.46	4.67	4.72	4.30	3.46	4.18	3.55	3.92	1.68
Mean	3.60	3.65	4.67	4.43	4.27	4.02	3.81	3.53	4.00	—
Range	.85	.32	.01	.87	.58	1.28	.92	.04	—	—

Table 2
ANALYSIS OF VARIANCE OF HOUSEWIVES' RATINGS
OF EIGHT NEW PRODUCT CONCEPTS

Source of variation	d.f.	Sum of squares	Mean square
Six-point hedonic scale, weighted 6, 5, 4, 3, 2, 1			
Among concepts	7	691.69	98.81 ^a
Among copywriters	2	30.39	15.19 ^a
Interaction	14	490.69	35.05 ^a
Residual	9,576	16,691.72	1.75
Total	9,599	17,904.50	—
Six-point hedonic scale, weighted 8, 4, 2, 1, 1, 1			
Among concepts	7	1,498.03	214.00 ^a
Among copywriters	2	94.36	47.18 ^a
Interaction	14	917.26	65.52 ^a
Residual	9,576	53,779.28	5.62
Total	9,599	56,288.93	—

^a Significant at the .01 level.

the basis of concept definitions outlined by the product's marketing staff.

Average ratings are shown for each concept and each copywriter's execution in Table 1. The data are shown in two ways: first, giving equal weight to each point on the six-point scale (simply 6, 5, 4, 3, 2, 1), and second, giving the weights of 8, 4, 2, 1, 1, 1 to the six points of the scale. Either version of the data could be used in computations. The first way is traditional, arbitrarily using equal weighting.

We prefer the second form because there is considerable experimental work that suggests the value of unequal weighting. On the basis of past data [1] we made our best judgment as to what the weights should be, in this case 8, 4, 2, 1, 1, 1.

Our previous experimentation with hedonic scales showed the general usefulness of logarithmic (e.g., 8, 4, 2, 1, 1, 1) weighting in terms of relative probability of purchase. On the lower points of the scale, there is little or no probability of purchase, but with a higher rating the probability of purchase increases geometrically.

An analysis of variance was computed for both sets of data in Table 1, the equally and unequally weighted scale points. Both variance analyses are shown in Table 2. Again, our approach was to analyze the unequally weighted data, although we also show the results of the traditional approach.

Duncan multiple comparison tests for differences among the eight concepts, the three copywriters (Table 3), and the 24 executions (Table 4) were also performed. The Duncan test resolves the difficulty, occasionally encountered in analysis of variance, of obtaining a non-significant range and a significant *F*-value [1].

ANALYSIS OF DATA

In practical terms, interpretation of the experimental results is similar, whichever scale weighting system is

Table 3
DUNCAN MULTIPLE-RANGE TESTS OF EIGHT CONCEPTS AND THREE COPYWRITERS

Weighting of scale data		Means ranked from best to worst ^a					
		Concepts			Copywriters		
Extremely desirable	Not at all desirable	Concept number		Copywriter number			
6,5,4,3,2,1		Best	Worst	Best	Worst		
		3 4 5 6	7 2 8 1	2	1 3		
8,5,2,1,1,1		3 4 5 6	7 2 8 1	2	1 3		

* Connecting lines indicate nonsignificant differences with the Duncan test at the .05 level.

used. In general, Concept 3 was the best, with a weighted (8, 4, 2, 1, 1, 1) mean of 4.67. Copywriter 2 was best, with a weighted mean of 4.14.

Not surprisingly, the effectiveness of any concept execution depends on the copywriter, since different copywriters do better with different concepts. This is shown

Table 4
DUNCAN MULTIPLE-RANGE TEST OF
24 COPY EXECUTIONS^a

Copywriter-concept	Scale weighting		Copywriter-concept	Mean
	6, 5, 4, 3, 2, 1	Mean		
2-4	5.03		2-6	4.74
1-3	5.00		2-4	4.72
2-3	4.99		3-4	4.72
3-6	4.98		1-3	4.68
2-6	4.96		2-3	4.67
3-3	4.96		3-3	4.67
2-5	4.89		2-5	4.54
3-5	4.68		3-5	4.30
3-7	4.65		3-7	4.18
1-7	4.57		1-7	4.00
1-5	4.50		1-5	3.96
1-6	4.47		2-1	3.89
2-1	4.44		1-1	3.88
1-4	4.43		1-6	3.86
1-1	4.42		1-4	3.85
2-2	4.35		2-2	3.78
1-2	4.31		1-2	3.70
1-8	4.27		3-8	3.55
3-8	4.25		1-8	3.54
3-2	4.22		2-8	3.51
2-8	4.19		3-2	3.46
3-6	4.14		3-6	3.46
2-7	3.87		2-7	3.26
3-1	3.78		3-1	3.04

* 2-4 designates the execution done by Copywriter 2 on Concept 4. Connecting lines indicate nonsignificant differences at the .05 level.

by a highly significant interaction effect in the analysis of variance and can also be seen in the sharply differing ratings for a concept depending on which copywriter did the work.

The interaction of copywriter and concept leads us to the conclusion that most concept tests are not testing only the relative value of different concepts: they are actually comparing the relative value of different executions. We would argue that they could just as well be called copy tests as concept tests. In fact, in advertising we are rarely seeking the *generally* best concept. The goal is to arrive at the most effective execution, from whichever concept makes this possible.

An advertising decision based on the experiments would select the highest-rated execution. Thus, in effect, we are not making a test of the concepts alone, but rather a test of the executions that result from interaction of concepts and copywriters.

In the array of 24 executions, the combination of Concept 6 and Copywriter 2 yielded the best execution, although we cannot claim statistically significant superiority over the next seven best-rated executions. However the data were weighted, the same six executions appeared among the best group, with no significant differences among them. The test then screened 24 executions and reduced them to a subset of reasonably good ones. This is the main usefulness of this sort of test—screening out sets of lower rating alternatives. But it is not, strictly speaking, concept testing. It is execution testing.

REFERENCES

1. Duncan, D. B. "Multiple Range and Multiple F Tests," *Technical Report No. 6*, Blacksburg, Va.: Virginia Polytechnic Institute, September 1953.
2. Haley, Russell I. and Ronald Gatty. "Adapting Attitude Measurement to Computer Processing," *Computer Operations: A Journal for Marketing Management*, 2 (May 1968), 11-6.

Copyright of *Journal of Marketing Research* (JMR) is the property of American Marketing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.