The Effects of Feature and Benefit Sentences in Advertising Copy on Consumers' Memory and Attitudes

Robert W. Meeds[1, *], Olan F. Farnall[2]

[1]Department of Communications, California State University, Fullerton, USA

[2]Department of Advertising, Texas Tech University, Lubbock, USA

Email address:
rmeeds@fullerton.edu (R. W. Meeds), olan.farnall@ttu.edu (O. F. Farnall)
[*]Corresponding author

To cite this article:
Robert W. Meeds, Olan F. Farnall. The Effects of Feature and Benefit Sentences in Advertising Copy on Consumers' Memory and Attitudes. *International Journal of Business and Economics Research*. Vol. 7, No. 5, 2018, pp. 136-143. doi: 10.11648/j.ijber.20180705.12

Abstract: The objective of this study was to empirically test how these dependent variables are influenced by features and benefits in advertising copy. Advertising copy testing research has produced conflicting results regarding the independence of recall, recognition and attitude constructs. Most copy-testing research, however, has focused on consumers' responses to the ad as a whole. The study reported here focuses on sentence-level manipulations to determine how variations in sentence patterns influence consumer memory and attitudes for advertised claims. A within-subjects experiment using magazine ads as stimuli was conducted in which the presence or absence of pairs of consumer benefit and product feature sentences were manipulated. The dependent variables were phrase recognition, morpheme recall and attitude toward the ad. Results showed that benefit and feature sentences did not improve recognition memory for phrases. However, ads with benefits or features enhanced morphemic recall and attitudes compared to the control condition. In particular, for ads that contained integrated benefit sentences and feature sentences together, an interactive effect was observed. Readers of these copy blocks displayed more positive attitudes toward the ad and higher levels of morpheme recall than benefits-only or features-only ads. The results are consistent with predictions from construal level theory, which has shown that benefit-based appeals are more effective in high construal situations while benefit- and attribute-based appeals are equally effective in low construal situations.

Keywords: Advertising, Copy Testing, Benefits, Features, Phrase Recognition, Morpheme Recall

1. Introduction

The purpose of this study is to investigate how consumer benefits and product features in print advertising influence readers' memory for ad copy elements and their evaluations of the ads. As concepts, features and benefits have long been linked together in the marketing, advertising and sales fields, with the general connection that customers may be impressed by a product's features but they are sold when they understand its benefits. Though the assumption has been with us for a long time, empirical research in advertising testing the idea that features and benefits work together is rare.

2. Literature Review

In a 1992 *Journal of Advertising* article on the future of advertising research, Stewart proposed four questions that "must be addressed if advertising is to continue to grow" (p. 1) [25]. The first question asked simply whether advertising worked, and if so, under what conditions. Studies published just before the Stewart article reported dismal results with regard to advertising's influence on sales, a staple measure of effectiveness at that time. For example, Abraham and Lodish [1] reported that only 49 percent of the 300 ads they examined produced an increase in respective product sales while Drane [8] reported little or no effect on sales as a result of advertising for almost 75 percent of the products in his sample. Perhaps that is why instead of accepting those numbers as a denigration of advertising researchers have focused on the second part of Stewart's question, the executional and contextual factors that impact advertising effectiveness.

Copy testing has long been the basic method advertisers

used to determine whether elements of an ad, or the ad in total, might succeed in getting attention, communicating an idea or properly positioning a brand. Even as early as 1975 estimates of advertising dollars devoted to copy testing were as high as $75 million [6]. Although most of that amount was not spent on academic research a number of published articles did emerge during that period, and the 20 years that followed, that established several important practices. For example, during that period researchers found that testing rough art versus finished work yielded comparable results for TV ads [4]. Another long-held assumption supported by copy testing methods of the time was that in persuasive communication simple language would result in higher memory scores [20, 23, 12]. It's the KISS (keep it simple stupid) concept still being taught to most beginning copywriters even though there is some evidence that copy complexity and ad readership are positively correlated [7]. A third assumption, and the focus of this study, was that effective print advertising used both features and benefits statements to score well on copy tests. Features delivered the product information while benefits statements provided the WHIIFM (what's in it for me).

As a method, copy testing was not without its critics. Opinions on the validity of copy testing have continued to vary even after the 1991 ARF Copy Research Validity Project. A frequently debated topic in copy testing research has been the relative merits (or lack thereof) of testing recall versus recognition memory.

However, the ARF project did reinforce the idea that copy testing works and that copy differences can be important. It suggested that ad likability might hold promise as well as multi-attribute testing but it did not provide an answer to Stewart's question of why one ad works and another does not [12]. One of the aims of this study is to explore the relationships between the common copy testing dependent variables recognition, recall and attitude toward the ad.

2.1. Sentence Level Analysis

Most copy testing techniques are designed to assess and compare audience reactions to entire messages. For example, researchers often use Starch scores to test the effectiveness of different ads (an aided recall measure in which consumers read a publication and then are asked which ads they noticed or read most of). This technique is also applied to testing the effects of different executional factors. Huhmann, et al examined the relationships between several advertising executional factors and Starch readership scores and found that copy block length was most negatively associated with an informational executional style and long copy blocks [14]. But accepted models of cognitive processing suggest people process advertising messages at varying levels depending on a host of antecedent conditions and confounding variables including motivation and ability to process [18]. To better understand of the complex processes involved in predicting how consumers respond to specific arguments contained in ad copy Meeds and Bradley suggest the sentence or phrase is the appropriate unit of analysis because it is on the same level as the argument.

"Applied to persuasive marketing or advertising texts, the sentence is the place where a product's features may be described or embellished, where a benefit to a consumer may be explicitly stated or implied, where the reader's curiosity piqued by the ambiguity of a visual pun is rewarded, and where emotions or themes are linked to a mental representation of a brand." (pp 104-105) [21]

They go on to suggest that in the sentence framework, cognitive processing can be influenced by average sentence length and arrangement of text surrounding the most important product features and benefits. The implication is that copy at the sentence level might be constructed in a manner to assist encoding and comprehension.

A few other studies have used sentence-level analysis but have provided contradictory findings with regard to the connection between syntax and persuasion. Meyvis and Janiszewski inserted irrelevant information in product description copy and found that the additional information weakens consumer's confidence that the product described could provide the benefit promised, a dilution effect [22]. Bradley and Meeds varied technical copy in print ads for consumer products and found that within limits, complexity of the ad had no significant negative effects on comprehension and negatively impacted attitude measures only at extreme overload situations [3].

In terms of more recent research on the impact of sentence level changes to body copy there are only a few studies to consider. The lack of direct examination of how features and benefits work together to produce impact was a bit troubling considering the fact that many of these long-accepted truisms are actually being taught to advertising students. But expansion into areas of psycholinguistics did present additional work. Using a sample of 480 brands whose print ads had been copy tested Lowrey, Shrum and Dubitsky were able to use linguistic feature analysis to connect three linguistic variables to increased memory for brand names, particularly in situations where the brand name was less familiar [17]. In a study that focused on TV commercials, Lowrey tested script complexity as a moderating factor for recall. She found that a complex message had a negative impact on recall and recognition [16]. These studies suggested that considering linguistic characteristics might be of particular benefit in developing brand names for new products. Clancy and Rabino used attribute/benefit desirability statements and found a high correlation between those statements and brand perceptions [5]. Samuelsen and Olsen tested functional claims based on tangible attributes and benefits against experiential claims focused on the promise of an experience [24]. They found that functional benefit claims outperform experiential claims in high involvement situations. Mayzlin and Shin examined attribute-focused appeals versus an appeal made with no information on the product attributes and found that the no attribute condition could cause consumers to engage in product search behavior. The combination of advertising with no attribute construction combined with search was rated as

much of a signal of quality as the product attribute condition [19].

Although these recent investigations provide useful information with regard to persuasion and psycholinguistic studies, it is surprising that no recent study has actually tested the sentence level impact of copy containing features, benefits or irrelevant information.

2.2. Recognition Versus Recall

Both scholars and practitioners have grappled for decades trying to sort out whether recognition and recall are better suited for measuring differences in degree (i.e., accessibility of memories) or for determining differences in kind (i.e., different encoding processes). Back in 1977, Krugman's classic article, "Memory Without Recall, Exposure Without Perception," was influenced by new advances in hemispheric brain research and he argued that high-involvement, rational processing was a left-brain task and low-involvement, emotional processing was a right brain task. He further argued this meant that recall was the most appropriate memory measure for print advertising while recognition was more appropriate for TV commercials [14]. Though some of Krugman's conclusions are now viewed as antiquated, recall and recognition measures are still important, but largely separate, pieces in the advertising effectiveness research puzzle, and remain largely detached from persuasion measures such as ad likeability and consumer attitudes [9]. And not surprisingly, our literature is rife with contradictory findings on these two measures. Some studies still support the distinctiveness of recognition and recall measures as indicators of different types of processing. For example, Leigh, Zinkhan and Swaminathan found recall to be influenced by cognitive aspects of print ads while recognition was influenced by affective factors [15].

The psycholinguistic and neuroscience literature has not reached a clear consensus on the relationships and differences between recall and recognition memory. The more prominent views, though, treat recall and recognition as related processes with some distinctive features [11, 13]. Similarly, in a structural equation model designed to test the relationship between recall and recognition memory for print ads, Finn found the two measures to be highly correlated but statistically distinct [10]. Based on this view, we offer separate but similar hypotheses for recognition and recall:

H1: Features and benefits will improve readers' recognition memory for ad copy.

H1a: Readers' recognition for ad copy will be higher in ads with benefit sentences compared to the control condition.

H1b: Readers' recognition for ad copy will be higher in ads with feature sentences compared to the control condition.

H1c: Benefits and features will interact so that readers' recognition for ad copy will be higher in ads with combined benefits and features than in ads with benefits or features alone or than the control condition.

H2: Features and benefits will improve readers' recall memory for ad copy.

H2a: Readers' recall for ad copy will be higher in ads with benefit sentences compared to the control condition.

H2b: Readers' recall for ad copy will be higher in ads with feature sentences compared to the control condition.

H2c: Benefits and features will interact so that readers' recall for ad copy will be higher in ads with combined benefits and features than in ads with benefits or features alone or than the control condition.

Though the relationships between memory, and attitudes are also frequently disputed in the advertising literature and there is good reason to view these common dependent variables as independent (i.e., remembering an ad does not equate with liking it), the effects of knowledge gain and comprehension on attitudes are held in more accord. Similar to Anderson's Information Integration Theory, if consumers are presented with advertising copy that lets them know how a product performs or what it does for them, this information is likely to be evaluated positively and it is likely to be incorporated into their overall attitudes toward the advertised brand [2]. These ideas are similar to the findings of Samuelsen and Olsen [24] and Mayzlin and Shin [19], and provide the basis for the third hypothesis, which is that the value of benefit or feature-related information will influence readers to react more positively to ads:

H3: Features and benefits will improve readers' attitudes toward the ad.

H3a: $A_{(ad)}$ will be higher in ads with benefit sentences compared to the control condition.

H3b: $A_{(ad)}$ will be higher in ads with feature sentences compared to the control condition.

H3c: Benefits and features will interact so that $A_{(ad)}$ will be higher in ads with combined benefits and features than in ads with benefits or features alone or than the control condition.

3. Methodology

3.1. Participants

Sixty-seven students (42 women and 25 men) enrolled in lower level mass communication classes at a large Southwestern university volunteered to participate. They received extra credit in a class for their participation.

3.2. Materials

The stimulus materials comprised eight black-and-white single page advertisements for fictitious brands in product categories deemed to be of interest to college students. These were *Collegiate Underwriters* (a renter's insurance firm), *Custom Phone Cards* (an international calling card service), *Dentabrite Tooth Whitening Strips*, *Dermaclear* (an over-the-counter acne treatment), *Evpure Pitcher Pro* (a drinking water purifier), *Soundmax* (an MP3 player), *Superior Standard Loans* (a student loan marketer), and *Ventarin Air Purifier*. Each ad had four versions—a zero benefits/zero features condition, a zero benefits/two features condition, a two benefits/zero features condition, and a two benefits/two features condition, for a total of 32 stimulus ads.

Although the ad layouts varied somewhat between the

brands–different typefaces, logo styles, and a combination of symmetrical and assymetrical layouts were used—each ad had a large illustration at the top (approximately half of the space), followed by a headline, a copy block and a logo. The amount of space devoted to the copy blocks was similar (approximately one-quarter of the space) across the eight fictitious brands. The target word length for the copy blocks was 100 words. In the two benefits/two features condition for example, the average number of words across the eight brands was 98 (range = 90-109). In addition, the structure of the copy blocks was controlled. Each ad began with an introductory paragraph, which we operationalize as thematic content. The second paragraph began with an introductory sentence (also thematic content), followed by the manipulation passage which consisted of a product feature sentence followed by a related consumer benefit sentence, and then a second feature sentence followed by a second benefit sentence. In the conditions with either zero benefits or zero features, these sentences were replaced by two filler sentences (e.g., "Acne can surface anywhere, although it usually appears on the face, back or chest."). After this section, the copy blocks finished with a concluding thematic section. In the condition with both zero features and zero benefits, only the two filler sentences appeared between the beginning and ending sections of thematic content. The headlines and thematic content were constant across treatment conditions.

3.3. Design and Procedure

The experimental design was a two (consumer benefit sentences present/consumer benefit sentences absent) by two (product feature sentences present/product feature sentences absent) within-subjects factorial design. Each participant read eight ads, two in each experimental condition. Data from the two observations in each condition were averaged prior to analyses.

Participants took part in the study in small groups of five to thirteen. The experiment took place in a research lab where each participant was seated at a computer. There was a partition between each participant to minimize distraction. Each participant was randomly assigned to one of 16 counterbalanced treatment orders. Post hoc tests of the dependent variables for all eight ads across the 16 orders yielded no significant order effects. The ads and the data collection measures were presented on the computer screens using MediaLab software. Following an introduction, instructions, and signing informed consent papers, the study proceeded. First, the participants read some on-screen instructions, and completed a few practice responses. Next came the stimulus presentation, in which the participants read the eight ads, each of which was on the screen for one minute. A pretest had shown an average comfortable reading time and ad viewing to be about 40 seconds per stimulus ad (s.d. = 7). Because the participants were informed they were going to be asked questions about the ads they had read, and because we wanted the exposure time to be the same for all participants, we did not want participants to feel rushed and

opted for a conservative exposure time of 60 seconds.

Next came a distraction task in which the participants were given a pencil and paper that contained a word search puzzle that they were asked to work on for ten minutes. After ten minutes a beep from their computers alerted them that the on-screen portion of the study was about to resume.

After the distraction task, attitude toward the ad *(Aad)* scale items were presented on the screen for each ad in the order they had been read, followed by free recall memory measures and recognition memory measures.

3.4. Measures

The *Aad* scales consisted of five, seven-point semantic differentials anchored by bad/good, not likeable/likeable, uninformative/informative, dull/interesting, and unclear/clear. Cronbach's Alphas across the eight ads ranged from .88 to .93, indicating the *Aad* scales were internally consistent. The mean *Aad* measures ranged from 4.00 to 4.67, indicating the ads were of average likeability.

Next, the recall measures asked participants to recall as many main words or portions of main words as they could for each ad. They were prompted with the fictitious brand name and given one minute for each ad. They were instructed to type each main word (or portion thereof) they could remember from the ad copy they had previously read and to not focus on words like articles, conjunctions and prepositions. They were instructed to press the enter key after each word they typed, after which a new blank line would appear. Participants were allowed up to 20 entries. Recall items were measured at the morpheme level. In coding the recall items, articles, conjunctions and prepositions were not counted. Remaining words or parts of words (e.g., if the participant typed "cover" when the actual word was "coverage") were coded as to whether they appeared in a benefit sentence, a feature sentence, in the headline or thematic sentence, or in a filler sentence. If a typed word actually appeared in two types of sentences, a partial score was given to each one. Additionally, words that participants typed but which did not actually appear in the ad copy were coded as guesses, and if any of the words the participants typed contained part of the fictitious brand name they were given credit as recalling the brand name. It is worth noting that participants were not asked to recall the brand name and were in fact given the brand name as a category cue at the beginning of the task, but we decided to code this item as a rough measure of brand name salience. Words that participants typed but that did not fit any of the above categories were coded as incorrect guesses. One measure was coded only in the two benefits/two features condition. For these ads, if a participant recalled at least one benefit word and one feature word, we coded how many pairs of the benefit and feature words recalled were from paired sentences. An additional measure, total morphemes recalled, was constructed by adding the number of morphemes from the benefit, feature, headline/theme and filler categories, and subtracting the number of incorrect guesses.

The next step was the recognition memory test. Here, ad-

related phrases appeared on the screen for two seconds. After each two-second presentation, the phrase disappeared and a screen with a radio button for "yes" and one for "no" appeared for two seconds. The participants' task was to click "yes" if they believed they had seen the words when they read the advertising copy earlier and "no" if they had not. The screen went blank after two seconds regardless of whether the participant had made a selection. Then the computer would beep to alert the participants to look at the screen again for the next set of words. The order of the items was the same as the ad presentation and the name of the fictitious brand appeared in the corner of the screen to remind them which ad they were being tested on. There were eight test items per ad presented in a random order—two were phrases taken from the two benefit sentences, two were from the two feature sentences, two were from thematic sentences and two were foils which had not appeared in any of the four ad versions. No items from the filler sentences were included in the recognition task. For example, the recognition items for the *DermaClear Acne Cream* ad were: "infused with Retin-A" *(feature)*, "dermatologist recommended" *(foil)*, "rapid cell turnover" *(foil)*, "contains no harsh chemicals" *(foil)*, "prevent new pimples" *(benefit)*, "your skin will heal" *(benefit)*, "doesn't magically go away" *(theme)*, and "get the clear picture" *(theme)*. Thus, if a participant read the two benefits/two features version of the *DermaClear* ad, a perfect score would have been six "yes" decisions on the feature, benefit and theme items and two "no" decisions on the foil items. Conversely, for a participant who read the zero benefits/zero features version, a perfect score would have been two "yes" decisions on the theme items and six "no" decisions on the benefits, features and foils items. In addition to coding the recognition data as the number of correct responses for benefits, features, headline/thematic content, and filler content, we then used a signal detection approach to calculate an overall score on each ad for hits, misses, false alarms and correct rejections. Hits and false alarms were then transformed into hit and false alarm ratio, from which a d' score was calculated using the equation d' = hit ratio – false alarm ratio.

4. Results

Hypothesis 1 predicted that the presence of features and benefits would improve readers' recognition memory for ad copy. A repeated measures ANOVA was run with the four treatment conditions of *benefits only, features only, benefits plus features,* and *no benefits/no features (control)* as the within-subjects factor. Participants' d' scores were the dependent variable. Mauchly's test for sphericity was significant (W = .64, X^2 = 26.2, p <.001), so the Greenhouse-Geisser adjustment for unequal variance was used in interpreting the F test. The omnibus F test was significant ($F_{(2.3, 136)}$ = 14.35, p <.001, η^2 = .20), indicating a significant difference among the four means. A test for linear contrasts with *benefits plus features* entered first, *benefits only* entered second, *features only* entered next and *control* entered last

was also significant ($F_{(1, 59)}$ = 16.41, p <.001, η^2 = .22). Pairwise tests, however, were not consistent with the predictions of hypothesis 1. H1a predicted that recognition would be better in ads with benefit sentences than in the control condition. The mean d' scores in the *benefits only* condition were 0.49 compared to 0.74 in the control condition. The differences were in fact significant in the opposite direction from what was predicted ($t_{(59)}$ = -4.38, p <.001). The same pattern held for H1b which predicted that recognition scores would be higher in the *features only* condition compared to the *control* condition (mean $_{(features)}$ = 0.46/ mean $_{(control)}$ = .74, $t_{(59)}$ = -4.87, p <.001). H1c predicted that recognition memory in the *benefits plus features* condition would be higher than in any of the other three conditions. This prediction was also not supported. The mean d' of 0.48 for the *benefits plus features* condition was not lower than the *benefits only* condition ($t_{(59)}$ = -0.38, p = .71), the *features only* condition ($t_{(59)}$ = 0.45, p = .66), or the *control* condition ($t_{(59)}$ = -4.40, p <.001). None of the tests for hypothesis 1 were supported (see Table 1).

Table 1. Recognition Scores (Hit Ratio - False Alarm Ratio).

Condition	Mean	SD	95% Conf. Int.	
			Lower	Upper
Benefits + Features	0.48	0.30	0.41	0.56
Benefits	0.49	0.22	0.44	0.55
Features	0.46	0.28	0.39	0.54
Control	0.74	0.38	0.65	0.84

Hypothesis 2 predicted that the presence of features and benefits would improve readers' recall memory for ad copy. Here, the same model repeated measures ANOVA as in hypothesis 1 was used, but with total number of morphemes recalled as the dependent variable. Mauchly's test for sphericity was not significant (W = .98, X^2 = 1.18, p = .95). The omnibus F test was significant ($F_{(3, 198)}$ = 6.17, p <.001, η^2 = .09), indicating a significant difference among the four means. A contrast test with *benefits plus features* entered first, *benefits only* entered second, *features only* entered next and *control* entered last was also significant and indicated a linear contrast as the best fit ($F_{(1, 66)}$ = 19.32, p <.001, η^2 = .23). A pairwise comparison showed the prediction from H2a that recall in the *benefits only* condition (mean $_{(benefits)}$ = 2.52) would be greater than in the *control* condition (mean $_{(control)}$ = 1.98) was supported ($t_{(66)}$ = 1.99, p <.05). For the H2b test, the mean for the *features only* condition (mean $_{(features)}$ = 2.41) was not greater than the mean for the *control* condition (mean $_{(control)}$ = 1.98, $t_{(66)}$ = 1.50, p = .07). H2c predicted that benefits and features together would result in the highest recall scores. Here, the mean for the *benefits plus features* condition was 3.13, which was higher than the *benefits only* condition (mean $_{(benefits)}$ = 2.52, $t_{(66)}$ = 2.34, p <.05), the *features only* condition (mean $_{(features)}$ = 2.41, $t_{(66)}$ = 2.66, p <.01), and the *control* condition (mean $_{(control)}$ = 1.98, $t_{(66)}$ = 4.52, p <.001). Thus, the predictions for recall memory dealing with benefits (H2a and H2c) were supported, but the prediction for features enhancing recall memory was not (see

Table 2).

Table 2. Recall Scores (Total Number of Morphemes minus Guesses).

Condition	Mean	SD	95% Conf. Int.	
			Lower	Upper
Benefits + Features	3.13	2.15	2.61	3.67
Benefits	2.52	2.09	2.00	3.02
Features	2.41	2.13	1.89	2.93
Control	1.98	1.99	1.49	2.46

Some additional analyses related to hypothesis 2 were run to investigate whether the types of morphemes being recalled varied with the presence of benefits and features. Looking at benefit-related morphemes in isolation, the number of benefit morphemes recalled in the *benefits plus features* condition (mean = 1.38) was not different from the *benefits only* condition (mean = 1.30, $t_{(66)}$ = 0.57, p <.57), nor was the average number of feature morphemes different between the *benefits plus features* condition (mean = 1.07) and the *features only* condition mean = 1.06, $t_{(66)}$ = 0.09, p <.93). In looking at *guesses* (morphemes recalled by participants that did not appear in the ad copy), there was a significant main effect for treatment condition ($F_{(3, 198)}$ = 4.73, p <.01, η^2 = .07). Mauchly's test for sphericity was not significant (W = .85, X^2 = 10.40, p = .06). A significant linear trend also occurred here ($F_{(1, 66)}$ = 16.57, p <.001, η^2 = .20), with the mean number of guesses of 0.88 in the *benefits plus features* condition being significantly lower than in the *benefits only* condition (mean = 1.16, $t_{(66)}$ = -2.15, p <.05), the *features only* condition (mean = 1.19, $t_{(66)}$ = -2.62, p <.01), and *control* condition (mean = 1.35, $t_{(66)}$ = -3.74, p <.001). Additionally, there was no difference in guessing in the *benefits only* condition (mean = 1.16) than in the *control* condition (mean = 1.35, $t_{(66)}$ = -1.79, p <.08), or between the *features only* (mean = 1.19) and the *control* (mean = 1.35, $t_{(66)}$ = -1.14, p = .26). (See Table 3).

Table 3. Guesses.

Condition	Mean	SD	95% Conf. Int.	
			Lower	Upper
Benefits + Features	0.88	0.74	0.70	1.06
Benefits	1.16	0.95	0.92	1.39
Features	1.19	0.91	0.96	1.41
Control	1.35	0.99	1.11	1.59

In looking at recall rates for headline and theme morphemes, the same pattern occurred as with the guesses. There was a significant main effect for treatment condition ($F_{(3, 198)}$ = 4.31, p <.01, η^2 = .06). Mauchly's test for sphericity was not significant (W = .94, X^2 = 4.22, p = .52). The linear trend was also significant for headlines and themes ($F_{(1, 66)}$ = 10.19, p <.001, η^2 = .13). For headlines and themes, the *benefits plus features,* there were fewer headline and theme morphemes recalled (mean = 1.56) than in the *control* condition (mean = 2.22, $t_{(66)}$ = -3.16, p <.01), and there also were fewer headline and theme morphemes recalled in the *benefits only* condition (mean = 1.76) than in

the *control* condition (mean = 2.22, $t_{(66)}$ = -2.39, p = .01). (See Table 4).

Table 4. Headline/theme Morphemes Recalled.

Condition	Mean	SD	95% Conf. Int.	
			Lower	Upper
Benefits + Features	1.56	1.20	1.27	1.85
Benefits	1.76	1.12	1.49	2.04
Features	1.87	1.34	1.54	2.20
Control	2.22	1.53	1.84	2.59

Hypothesis 3 predicted that benefits and features would enhance attitudes toward the ad ($A_{(ad)}$). The same analysis approach used for the first two hypotheses was used here. The omnibus F test was significant ($F_{(3, 195)}$ = 9.82, p <.001, η^2 = .13), and Mauchly's test for unequal variances was not significant (W = .88, X^2 = 8.43, p = .13). As was the case with the first two hypotheses a significant linear trend was found as a within-subjects contrast when the *features plus benefits* condition was entered first, followed in order by the *benefits only, the features only,* and the *control* conditions ($F_{(1, 65)}$ = 15.70, p <.001, η^2 = .20). $A_{(ad)}$ in the *benefits only* condition (H2a, mean = 4.27) was higher than the *control* condition (mean = 3.97, $t_{(65)}$ = 2.58, p <.01). $A_{(ad)}$ was also higher in the *features only* condition (H2b, mean = 4.50) than in the *control* condition (mean = 3.97, $t_{(65)}$ = 3.68, p <.001). $A_{(ad)}$ was significantly higher as predicted in H3c in the *benefits plus features* condition (mean = 4.64) than in the *benefits only* condition (mean = 4.27, $t_{(65)}$ = 3.05, p <.001), and the *control* condition (mean = 3.97, $t_{(65)}$ = 4.47, p <.001), but not the *features only* condition (mean = 4.50, $t_{(65)}$ = 1.16, p <.13). Thus, hypothesis 3a and 3b were supported, and two of the three predictions in hypothesis 3c were supported. (See Table 5).

Table 5. Attitude Toward the Ad.

Condition	Mean	SD	95% Conf. Int.	
			Lower	Upper
Benefits + Features	4.64	1.14	4.36	4.92
Benefits	4.27	1.03	4.02	4.53
Features	4.50	1.10	4.23	4.77
Control	3.97	1.24	3.66	4.27

5. Discussion

The results for the recognition and recall measures observed in this study were inconsistent. Consistent with the hypotheses, recall memory for words in advertising copy improved when benefit or feature sentences were included. In particular, the total number of morphemes recalled was the highest in the *benefits plus features* condition, followed by the *benefits only* and *features only* conditions, with the *control* condition having the lowest recall. In addition to recalling more words overall, participants recalled fewer headline or theme morphemes and made fewer incorrect guesses when they read ads that had benefits or features or both. This suggests that the benefits and features content was

also more readily retrieved, although a direct comparison with the *control* condition was not possible. This is consistent with a construal levels theory approach, such as used by Hernandez, et al, who found that consumers with high construal levels found benefit-based appeals more persuasive while consumers with low construal levels found benefit and attribute-based appeals equally persuasive [13].

Conversely, recognition memory did not perform as predicted in hypothesis 1. In fact, recognition memory scores were highest in the *control* condition. Drawing the conclusion from this null result that recognition and recall operated as distinct processes is not warranted. It could be that the two measures were in fact distinct from each other in this case, or there could be other explanations for the null result in H1. For example, the recognition test in the *control* versions of the ads contained two targets and six foils, compared to six targets and two foils in the *benefits plus features* condition, and four of each in the *benefits only* and *features only* conditions. Recognition scores were low in general, regardless of treatment condition. It could be that, near the end of a 40-minute research session that involved lots of reading, participants simply struggled more with the task of identifying phrases they had already seen than they did with correctly rejecting phrases they had not. As Kosslyn noted, "hemispheric specialization will be understood not in terms of general dichotomies among types of materials or tasks … but in terms of the underlying structure of information processing" (p. 159) [13].

Also, making inferences about the efficacy of recall and recognition measures in other types of advertising memory tasks is not warranted. The bulk of the published advertising research combining recognition and recall measures has been focused on the ad as a whole. And although there can be verbal components to recognizing a whole ad (e.g., recognizing the headline in a print ad), we argue that ad recognition is primarily a visual task in most studies. And although many issues concerning hemispheric specialization are still hotly debated in the neurosciences, most psychologists agree that verbal-only encoding is a left-lateralized process while image-only (e.g., unfamiliar faces) is a right-lateralized process, and encoding information that has both visual and verbal components is a bilateral process. The differential results observed here are of most interest to advertising researchers engaged interested in lexical memory.

The $A_{(ad)}$ scores were also highest for the ads with benefit and feature sentences. *Benefits plus features* $A_{(ad)}$ was significantly higher than *benefits only* and $A_{(ad)}$ in all three treatment conditions was higher than in the *control* condition. This linear relationship supports the longstanding but rarely tested idea that benefits and features together are more persuasive.

In looking at the linear relationships of the dependent variables (see Figure 1), it is tempting to conclude that $A_{(ad)}$ and recall are positively correlated with each other and inversely correlated with recognition memory. Both statistically and visually, however these are aggregate results.

As a *post hoc* analysis, we examined correlations for $A_{(ad)}$, recall and recognition in all four conditions. Of the twelve correlation tests, only one was significant ($r = .32$ between $A_{(ad)}$ and recognition memory in the *features only condition*. None of the other correlations were significant or greater than ±.18. Although this does not establish independence among these three concepts conceptually, it does appear that attitude formation and the two types of memory operated independently in this study.
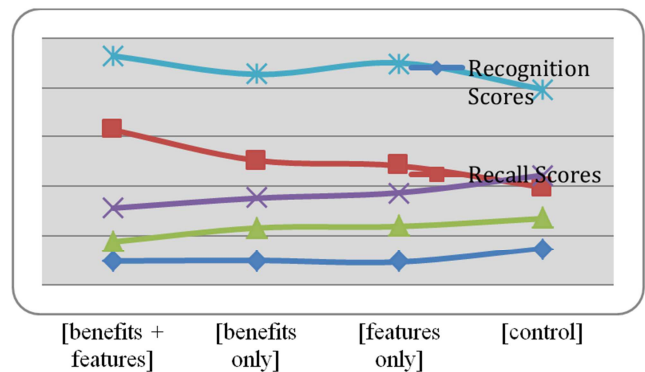

Figure 1. *Line charts for recognition, recall, guesses, headline/theme recall and attitude toward the ad across experiment conditions.*

6. Conclusion

The implications of this study on advertising practice are clear, though not novel. The results of this study provide evidence that readers respond positively and retain more product-related information when features and benefits are integrated with each other. Though this relationship has been presumed for decades, it has not been empirically confirmed in the advertising academic literature. The primary implication of this study for advertising scholars is that sentence-level manipulations were shown to independently influence both attitudes for ads and recall for specific types of words.

A limitation of this study is that stimulus materials were limited to print ads and to ads for products and services where a variety of features and benefits could be juxtaposed. The task demands of the recognition task may have also created a floor effect.

Future research should expand the use of sentence level language manipulations to non-print advertising environments and to ads using emotional appeals.

References

[1]   Abraham, Magid and Leonard Lodish (1989), Advertising Works, A Study of Advertising Effectiveness and the Resulting Strategic and Tactical Implications, Chicago: Information Resources, Inc.

[2]   Anderson, Norman H. and Cheryl C. Graesser (1976). An information integration analysis of attitude change in group discussion. Journal of Personality and Social Psychology, Vol 34 (2), Aug 1976, 210-222.

[3]   Bradley, S. D. & Meeds, R. (2004). "The effects of sentence-level context, prior word knowledge, and need for cognition on information processing of technical language in print ads," Journal of Consumer Psychology, 14 (3), 291-302.

[4]   Caffyn, John M. (1965), "Teleplex Testing of TV Commercials," Journal of Advertising Research 5 (2), 29-37.

[5]   Clancy, Kevin J, and Samuel Rabino (2007), "The Effects of Visual Enhancement on Attribute/Benefit Desirability and Brand Perception Measures: Implications for Reliability and Validity," Journal of Advertising Research 47 (1), 95-102.

[6]   Clancy, Kevin J, and Lyman E. Ostlund (1976), "Commercial Effectiveness Measures," Journal of Advertising Research 16 (1), 29-34.

[7]   Chamblee, Robert, Robert Gilmore, Gloria Thomas and Gary Soldow (1993), "When Copy Complexity Can Help Ad Readership," Journal of Advertising Research 33 (3), 9-21.

[8]   Drane, Robert, (1988), "Boosting the Odds of Advertising Success," in Evaluating the Effects of Consumer Advertising on Market Position Over Time: How to Tell Whether Advertising Works, Stephen Bell, ed., Summary of A Marketing Science Institute Conference, Cambridge, MA: Marketing Sciences Institute.

[9]   Du Plessis, Erik (2005). The Advertised Mind: Ground-Breaking Insights into How our Brains Respond to Advertising. New York: Millward Brown.

[10] Finn, Adam (1992). "Recall, Recognition and the Meausurement of Memory for Print Advertisements: A Reassessment," Marketing Science, 11: 1, 95-100.

[11] Haist, Frank, Arthur P. Shimamura, and Larry R. Squire (1992). "On the Relationship Between Recall and Recognition Memory," Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol 18 (4), Jul 1992, 691-702.

[12] Haley, Russell I. and Allan L. Baldinger (2000), "The ARF Copy Research Validity Project," Journal of Advertising Research 40 (6), 114-135.

[13] Hernandez, J. C., Wright, S. A., and Rodrigues, P. F. (2015). "Attributes versus benefits: The role of construal levels and appeal type on the persuasiveness of marketing messages." Journal of Advertising, 44 (3), 243-253.

[14] Huhmann, B. A. Franke, G. R., and Mothersbaugh, D. L (2012). "Print advertising: Executional factors and the RPB grid." Journal of Business Research, 65, 849-854.

[15] Kosslyn, Stephen M. (1987). "Seeing and Imagining in the Cerebral Hemispheres: A Computational Approach," Psychological Review, 94: 2, 148-175.

[16] Krugman, Herbert (1977, reprinted 2000). "Memory Without Recall, Exposure Without Perception," Journal of Advertising Research, 40: 49-54.

[17] Leigh, James H., George M. Zinkhan, and Vanitha Swaminathan (2006). "Dimensional Relationships of Recall and Recognition Measures with Selected Cognitive and Affective Aspects of Print Ads," Journal of Advertising, 35: 1 105-122.

[18] Lowrey, Tina M. (2006), "The Relation Between Script Complexity and Commercial Memorability," Journal of Advertising 35 (3), 7-15.

[19] Lowrey, Tina M., L. J. Schrum, and Tony M. Dubitsky (2003), "The Relation Between Brand-Name Linguistic Characteristics and Brand-Name Memory," Journal of Advertising 32 (3), 7-17.

[20] MacInnis, Deborah J. and Bernand J. Jaworski (1989), "Information Processing from Advertisements: Toward an Integrative Framework," Journal of Marketing 53 (4), 1-23.

[21] Mayzlin, Dina and Jiwoong Shin (2011), "Uninformative Advertising as an Invitation to Search," Marketing Science 30 (4), 666-685.

[22] Meeds, Robert (2004), "Cognitive and Attitudinal Effects of Technical Advertising Copy: The Roles of Gender, Self-assessed and Objective Consumer Knowledge," International Journal of Advertising, 23, 309-335.

[23] Meeds, Robert, and Bradley, S. (2007). "The role of important product attribute sentences in advertising recall and attitudes." In T. Lowrey (Ed.), Psycholinguistic Phenomena in Marketing Communications. Mahwah, NJ: Lawrence Erlbaum.

[24] Meyvis, Tom and Chris Janiszewski (2002), "Consumers' Beliefs about Product Benefits: The Effect of Obviously Irrelevant Product Information," Journal of Consumer Research 28 (3), 618-635.

[25] Ostlund, Lyman E. (2002), "Advertising Copy Testing" A Review of Current Practices, Problems and Prospects," Current Issues and Research in Advertising 24 (1), 87-105.

[26] Samuelsen, Bendik M. and Lars Erling Olsen (2010), "Promising Attributes and Experiences: Attitudinal Responses to Functional versus Experiential Ad Claims and the Moderating Role of Involvement," Journal of Advertising 39 (2), 65-77.

[27] Stewart, David W. (1992), "Speculations on the Future of Advertising Research," Journal of Advertising 21 (3), 1-18.