


Creative That Sells: How Advertising Execution Affects Sales

Nicole Hartnett, Rachel Kennedy, Byron Sharp & Luke Greenacre


To cite this article: Nicole Hartnett, Rachel Kennedy, Byron Sharp & Luke Greenacre (2015): Creative That Sells: How Advertising Execution Affects Sales, Journal of Advertising, DOI: [10.1080/00913367.2015.1077491](https://doi.org/10.1080/00913367.2015.1077491)

To link to this article: <http://dx.doi.org/10.1080/00913367.2015.1077491>

 View supplementary material 

 Published online: 29 Sep 2015.

 Submit your article to this journal 

 Article views: 235

 View related articles 

 View Crossmark data 

Creative That Sells: How Advertising Execution Affects Sales

Nicole Hartnett, Rachel Kennedy, Byron Sharp, and Luke Greenacre

Ehrenberg-Bass Institute for Marketing Science, University of South Australia, Adelaide, Australia

Advertising creative is widely accepted as critical to advertising success. However, generalizations of what works in applied settings across different conditions are few. The present study replicates the seminal work of Stewart and Furse (1986), who investigated the effect of more than 150 creative devices on several copy-testing measures of advertising effectiveness. We replicate the analysis using the original codebook but examine the link to in-market, short-term sales effectiveness. We use a large sample of 312 television ads from several product categories aired in multiple countries. Our findings indicate that the codebook remains relevant for characterizing current advertising practices but many of the creative devices found most (or least) effective differ to those from the original study. Similar to Stewart and Furse (1986), no single creative device can do much alone to explain sales effectiveness. There is no one simple cookbook for making sales effective advertising, though such research offers some important guidelines.

Advertising creative, or content, is commonly acknowledged as the primary determinant of advertising effectiveness. Many studies drawing on diverse methods, including sales data audits (Blair and Kuse 2004), split-cable experiments (Lodish et al. 1995), and single-source analysis (Wood 2009), show enormous variation in the sales effects of different creative; variation that is larger than that recorded for typical changes in advertising weight or expenditure. Wood (2009),

for example, found that powerful creative was 10 to 20 times more sales effective than mediocre creative.

Exploring the efficacy of the various creative message strategies (i.e., *what* is communicated) and creative execution devices (i.e., *how* it is communicated) used in advertising is therefore an important area of research that has a long history. Initial studies of advertising's creative effects examined mechanical devices (e.g., size, headline, color) that improved print advertising recognition and readership (Twedt 1952), which evolved to include message devices (e.g., quality appeals, comparisons) (Holbrook and Lehmann 1980), and then sound and motion devices for broadcast advertising (e.g., background music played, voiceovers, scene changes) (Stewart and Furse 1986).

How advertising creative contributes to advertising effectiveness continues to attract interest and debate in academic journals. Researchers have theorized and investigated how the dimensions of creativity (e.g., divergence, relevance) (Lehnert, Till, and Ospina 2014; Sasser and Koslow 2008; Smith and Yang 2004) or the nuances of specific creative devices relate to consumers' responses to advertising. Most recent examples include rhetorical figures (Theodorakis, Koritos, and Stathakopoulos 2014), gender stereotypes (Eisend, Plagemann, and Sollwedel 2014), and comedic violence (Yoon and Kim 2014). Exploring different conditions (e.g., products, services, social issues) and consumer groups (e.g., various demographic factors, involvement with the ad or brand) has also revealed differential responses, suggesting there are boundaries to some creative effects. These studies, along with many others, have made important contributions to developing the different theories of how advertising works, particularly in regard to how different consumers process and evaluate advertising.

One possible caveat to this vast body of research, however, is that most advertising creative research is conducted as laboratory experiments, which suffer many limitations. Experiments, which enable us to hold many variables constant, are a means to demonstrate causality given the complex phenomena that advertising creative represents. Researchers should, however, incorporate procedures that provide evidence of external validity to establish practical relevance. Unfortunately, many advertising experiments are far removed from real advertising situations because they test only a handful of ads for fictitious

Address correspondence to Nicole Hartnett, Ehrenberg-Bass Institute for Marketing Science, School of Marketing, University of South Australia, Level 4 Yungondi Building, City West Campus, North Terrace, Adelaide, Australia 5000. E-mail: Nicole.Hartnett@marketingscience.info

Nicole Hartnett (MBA (Res), University of South Australia) is a Senior Research Associate, Ehrenberg-Bass Institute for Marketing Science, University of South Australia.

Rachel Kennedy (PhD, University of South Australia) is Associate Director, Ehrenberg-Bass Institute for Marketing Science, and Associate Professor of Marketing, University of South Australia.

Byron Sharp (PhD, University of Adelaide) is the Director, Ehrenberg-Bass Institute for Marketing Science, and Professor of Marketing, University of South Australia.

Luke Greenacre (PhD, University of Technology Sydney) is a Senior Research Associate, Ehrenberg-Bass Institute for Marketing Science, University of South Australia.

or unfamiliar brands, force exposure on participants, and/or use intermediate measures of advertising effects (e.g., on memory, attitudes, intentions) (McQuarrie, Phillips, and Andrews 2012). Such intermediate measures are typically found to be poor predictors of actual behavior or in-market sales (Haley and Baldinger 1991; Lodish et al. 1995). It is often unclear, then, how effects measured under such conditions may manifest in real market settings, where the most important objective is typically sales.

Few studies have examined the effects of different advertising creative strategies or devices on in-market consumer behavior directly (e.g., Chandy et al. 2001), and hence generalizations across findings must be made cautiously (Tellis 2009). There is still much to be learned regarding which of the many creative devices used in advertising are more or less associated with sales. Importantly, agreement is broad that the key outcome of brand advertising is sales. This holds for researchers and professionals with varied theoretical backgrounds, from proponents of traditional hierarchy of effects theories through to newer low-involvement emotional theories (Heath 2009; Vakratsas and Ambler 1999).

The purpose of the present study is to replicate and extend the work of Stewart and colleagues (Stewart and Furse 1986; Stewart and Koslow 1989), who developed a comprehensive codebook of more than 150 creative devices that were linked to several intermediate measures of advertising effectiveness. We closely replicate the codebook and analyses with a new large sample of television ads; however, our dependent variable differs from those of the original studies. Where they used intermediate measures of advertising effects, we are concerned with a measure of short-term sales effect obtained from single-source data.

Single-source data, which continuously records product purchases and television advertising exposures for individual households, allows for a quasi-experimental design extracted from “as it lies” real market data. Purchases that are preceded by ad exposure are compared to purchases that are not; such is the case for Jones’s (1995) Short-Term Advertising Strength (STAS) measure. Researchers can also control for some confounding variables (e.g., advertising frequency, price promotions), while other environmental factors (e.g., distribution) are assumed constant across exposure groups, similar to a split-cable experiment. Single-source data therefore provides the least contaminated measure of the short-term impact of advertising creative on category buyers’ purchases.

Though most intermediate measures do not predict sales well, Research Systems Corporation (RSC) has demonstrated reasonably high correlations ($r = .60-.70$) between their ARS Persuasion measure used by Stewart and colleagues and short-term market share changes (Blair and Kuse 2004). The apparent closeness of the respective dependent variables provides a strong basis for replicating the findings across studies to establish generalizable patterns of creative effects across data sets and time. Our study is a replication with extension, with a new

and arguably more robust measure of sales in response to advertising.

REPLICATING STEWART AND FURSE (1986)

At the time lacking accepted theories of how advertising works, early studies sought to catalog the many different creative devices used by advertisers (e.g., McEwen and Leavitt 1976; Stewart and Furse 1986). Researchers reviewed the relevant literature and interviewed authorities in advertising research and creative fields. These studies developed reliable codes that could be consolidated into a few coherent categories, which accounted for substantive variability across ads. This was an important first step to (1) establish clear definitions of creative devices, accepted by both researchers and practitioners, and (2) provide initial findings that could be modified and extended to new contexts over time.

Stewart and Furse’s (1986) study is regarded as the seminal work because of its unprecedented scope. The Marketing Science Institute sponsored the study, and RSC provided a sample of 1,059 ads for 356 brands of mostly consumer packaged goods (CPGs) products, almost half (44%) of which were for new products. From 1981 to 1983, RSC copy-tested the ads in a forced exposure laboratory setting for their clients, and from these tests provided three measures of advertising effects: related recall (i.e., percentage of people who claimed to have seen/could play back unique details of the ad three days after exposure), message comprehension (i.e., percentage of people who could play back the intended message), and a pre-post “persuasion” measure (i.e., percentage of people who chose the brand as part of a lottery draw immediately after exposure minus the percentage who chose the brand before exposure, adjusted for the expected level of persuasion without exposure). Stewart and Furse (1986) examined the effect of more than 150 creative devices on each of the measures and compared effects for new and established products. It was the first study of its kind to (1) consider many creative devices with concern for establish their codebook’s reliability, (2) analyze a large sample of television ads, and (3) use several validated measures of advertising effects.

The *Journal of Advertising Research (JAR)* first published a “tease” of results (Stewart and Furse 1984), and *Psychology and Marketing* then published a more comprehensive paper (Stewart and Furse 1985). The full results were published as a book (Stewart and Furse 1986). *JAR* also reprinted the original 1984 paper in 2000, in a special issue celebrating articles that made contributions of lasting value. Later, the *Journal of Advertising (JA)* published a replication of the original study (Stewart and Koslow 1989). It was an exact replication, with another large sample of 1,017 new ads, also copy-tested by RSC from 1980 to 1987. Again, many of the ads were for new products (40%).

The results showed some consistencies between the two studies:

- Each of the copy-test measures was positively though weakly correlated; recall and comprehension were the most correlated.
- Brand-differentiating messages, which claim that something about the product is unique to the advertised brand, was the most important determinant of recall and persuasion, though it explained only a small percentage of the variance of any of the effects measures.
- Some other consistent findings included, for example, nutritional information had a strong negative relationship with all measures, while humor and the amount of time the actual product is on screen had strong positive relationships with recall. These explained very little of the variance (about 1%) of any of the effects measures.

These two studies have been rather influential. Together they had almost 600 citations according to Google Scholar as of January 2015. Others have used the findings to develop guidelines for advertising best practice, such as Armstrong's (2010) persuasion principles. The Stewart and Koslow (1989) replication also won the *JA*'s Best Article award that year, all of which speaks to the importance of these early works.

In the decades that followed, several studies partially replicated the codebook. [Stanton and Burke \(1998\)](#) and [Phillips and Stanton \(2004\)](#), for example, replicated 24 of Stewart and Furse's (1986) creative devices to examine effects across different conditions, including 15- and 30-second ads and young and mature consumers. These studies also used large samples of ads copy-tested by RSC and, not surprisingly, found results that concur with Stewart and colleagues. By drawing from overlapping samples of ads and using the same measures, these studies provide limited evidence to the generalizability of the original findings.

There has been a recent extension into a newer context by Bellman, Schweda, and Varan (2012), who replicated 16 creative devices and examined the effects of these on response rates to interactive television (iTV) ads in the United Kingdom. Only background cast had a positive main effect on response rates, which contradicts the findings of Stewart and colleagues, who reported consistent negative effects on comprehension and persuasion measures. Even so, the strength of this particular study is its real market data, using a behavioral dependent variable. Behavioral or sales-based measures have long been the gold standard for advertising research, but the norm has been to accept intermediate measures because they are easier to collect. Behavioral measures have become increasingly sought after, however, as businesses demand more financial accountability from marketing expenditures ([Stewart 2009](#)) coupled with technological advances that provide better measures of larger samples over longer periods of time. We commend Bellman, Schweda, and Varan (2012) but note that iTV ads are a rather unique context, where responses are immediate and pertain to requests for information or free samples rather than product purchases. Much advertising has a

delay of days, weeks, or even months before consumers can act on what they have seen or heard. What stimulates clicks for iTV ads may not generalize to more traditional advertising and its affect on sales.

The objectives for the present study were twofold. First, to replicate the Stewart and Furse (1986) codebook in its entirety on a large sample of contemporary television ads. Replication allows for the evaluation of the codebook's continued relevance when examining how advertising is executed. Second, to determine whether the original findings related to the creative devices that most affect advertising performance generalize to our new data set of ads and to a more relevant measure of advertising effectiveness.

METHOD

Sample of Ads

A global CPGs manufacturer provided us with 312 ads that were aired between 2000 and 2013. The ads are for more than 60 brands in four categories (two staples and two impulse) across five markets, including the United Kingdom (33%), Germany (28%), France (22%), the United States (17%), and Australia (<1%). Of these, 259 ads are for brands owned by the manufacturer and 53 are for rival brands across categories. The brands vary in size, from established market leaders to small but stable brands. The data set includes a variety of ad lengths, but most ads are 30 seconds (48%) or 20 seconds (33%).

The present data set is different and somewhat less varied than Stewart and colleagues' data sets, which may affect the comparability of findings. Our ads are all for food products, mostly from a single firm and mostly from European markets, whereas the original studies included food, over-the-counter remedies, and personal care and household products for brands from dozens of firms, and all were from the United States. Different laws bind different countries, and these can also lead to differences in creative execution. For example, comparative advertising is legal and subsequently more common in the United States. Our data also included fewer ads for new products, 27% versus > 40%, which could reflect different proclivities for innovation across continents or that the original studies were of copy-tested commercials, which skewed toward ads for product launches. Differences between the data sets will be revisited during the results discussion.

Sales Effectiveness (Dependent Variable)

Our industry partner determined the sales effectiveness of the ads using single-source data. Each country had a local panel with a minimum of 5,000 households available to analyze. The effects were measured based on purchases scanned in a four-week period, with advertising exposures logged via set-top boxes in the four-weeks prior. The measure of short-term sales

effectiveness is an index that compares brand purchases made by exposed and unexposed households, similar to STAS (Jones 1995), but uses a complex set of contingency tables to account for other impacting and extraneous variables, such as the frequency of advertising exposure and promotional activity. Purchases made at a discount, for example, were excluded. The approach therefore estimates the probability of a brand purchase based on being exposed to the brand's advertising (versus not) such that the sales indexes isolate brand and product (SKU) sales changes driven by specific advertising executions alone.

The sales indexes differed by category and country (i.e., some categories/countries were more responsive to the advertising than others); hence, they were converted to a three-level ordinal variable of good, average, and bad sales effectiveness. The variable was empirically derived from a systematic review of historic data, which had been driving the company's advertising decision making prior to this study. The cutoffs for good and bad ads were ± 5 percentage index points from the average index score, giving confidence that the relative sales effectiveness of the outer groupings was meaningfully different, as well as consistent across categories, countries, and time periods. Of the 312 ads, 34% were bad, 29% were average, and 37% were good.

We checked for potential confounding relationships between the characteristics of the ads and sales effectiveness but found none. Chi-square tests found no significant relationships for sales with product category, $\chi^2(6) = 6.67, p = 0.35$, or country, $\chi^2(8) = 5.90, p = 0.66$, and one-way analysis of variance (ANOVA) found no significant differences for ad lengths, $F(5, 306) = 0.49, p = 0.78$.

Coding Procedure (Independent Variables)

The procedures for coding the content of the ads were conducted in line with Stewart and Furse (1986) except that five judges were used, rather than four judges, to more easily establish a majority for the binary creative devices that are simply either present or absent. The 42 coders were paid university students, and the main criterion for eligibility was fluency in a language portrayed by the ads: English, German, and/or French. Coders received the codebook in subsets based on the existing classifications (e.g., information, visual devices) to minimize coder fatigue by limiting coding sessions to three to four hours. Not all coders participated in all modules.

The codebook was kept the same except for two variables: number of words (D106) and visual pace (D107). These were changed from ordinal variables (above average, average, below average) to counts because it was unclear whether the average words and pace from the 1980s still applied. Also, any codes that referred to the "first 10 seconds" were changed to the "first third" to reflect the different ad lengths in our data. We provided coders with some updated examples relevant to the specific categories in our

data, for example, which manufacturer owns which brands. Coders were trained by one of the authors, who explained each code and walked through coding an example ad that was not from the data set. Coders were encouraged to ask questions to minimize ambiguity.

Intercoder reliability was assessed based on the average pairwise percent agreement (APPA) across coders. APPA was 86% across nominal codes, though most achieved more than 80% APPA—well above chance. Only codes above 60% APPA were included in our analysis, which exceeds the benchmarks suggested by Rust and Cooil's (1994) proportional reduction in loss reliability measure with five coders. Coding discrepancies (for nonbinary variables) were adjudicated by one of the authors.

RESULTS

How Often the Creative Devices Occurred

The most common tones were humor (29%), cute/adorable (16%), warm/caring (9%), and happy/fun-loving (9%), all positive. Most ads steered clear of negative messaging, focusing primarily on positive appeals (90%). Most ads also included some form of information (82%), such as ingredients (40%), news (27%), and/or sensory experiences (18%). Some ads, however, simply reminded consumers that the product exists rather than communicating specific attributes or benefits (25%). Many ads had a demonstration of the product in use (77%) in a setting directly related to that product's use (61%). Continuity of action (having an obvious beginning, middle, and end) was dramatically more common than vignettes (86% versus 4%), and most ads included on-screen principal characters (88%). Despite the high occurrence of on-screen characters, many ads delivered the audio message with a voiceover only (56%), without on-screen character dialogue. A full table of this descriptive analysis, as well as reliability results, is available as an online appendix.

The relative frequencies for nominal variables found in our data set are positively correlated with those reported by Stewart and Furse (1986), $r(142) = .59, p < .01$. The correlation suggests that ads from the past decade, as represented by our data set, use broadly similar characteristics to ads from the 1980s.

Unsurprisingly, there were some differences, for example, there were far fewer superiority claims (1% versus 73%), indirect comparisons (1% versus 28%), and brand-differentiating messages (1% versus 44%). Overall, these sorts of "hard-selling" tactics were very rare in our data, which does appear to suggest a move toward softer selling over time. However, these differences may instead reflect the differences in the data sets. Food products may be less inclined or able to claim superior or unique attributes or benefits than personal care and household products, or European advertisers may prefer subtler indirect sales claims compared to U.S. advertisers.

Such differences also meant that we could not repeat the main finding of Stewart and colleagues regarding brand-differentiating messages. Only three of our ads (i.e., 1% of the data set) included brand-differentiating messages, which is an insufficient sample to explore a meaningful relationship with sales. In line with Stewart and Furse (1986), only creative devices present in more than 5% of ads were included in the following analyses against the dependent variable.

Relationship Between Creative Devices and Sales Effectiveness

Prior to more complex analysis, chi-square and one-way ANOVA analyses were conducted to examine the association of each creative device with sales effectiveness.

Table 1 shows that only a few creative devices had a direct relationship with sales effectiveness. Creative devices not shown in this table had nonsignificant relationships with sales.

Psychological benefits, humorous tones, no principal characters, including voiceovers with characters that speak (rather than only voiceovers), and incorporating some negative appeals (negative experiences that the product helps avoid, rather than predominantly positive appeals) all had positive relationships with sales. When these creative devices were present there were more good ads, with fewer average and/or bad ads, than when they were absent. Music (of any kind), music that creates mood (actively tries to evoke emotions,

rather than background, elevator-style music), demonstrations of the product in use (e.g., cooking, consuming), and puffery (unsubstantiated superiority claims) all had negative relationships with sales.

Table 2 compares these results to the relationships ($p < .05$) found by Stewart and colleagues for the ARS persuasion measure specifically. The one consistent result between our study and Stewart and Furse (1986) is the positive effect of having no principal characters. Stewart and Furse (1986) believed that less time spent on characters meant more time devoted to product demonstrations, which also had a positive effect on persuasion. For our data, however, demonstrations (in use) were negatively related to sales; therefore, ads without principal characters were possibly more effective because they less often included demonstrations. Another notable result is that of humorous tones. Though Stewart and colleagues found no relationship with persuasion, humor had strong positive effects on recall and comprehension. They suggested that humor gets attention, but other factors, such as brand-differentiating messages, are also needed to motivate consumers to purchase the brand. We do, however, find a direct relationship to sales.

It is possible that in this analysis, the creative devices may be contributing to the effect or may be correlates of other creative devices present in the same ads. Regression analysis that controls for the effects of other creative devices can provide further insight into these relationships.

TABLE 1
Chi-Square Results for the Creative Devices Related to Sales Effectiveness

Creative Devices	χ^2	df	Freq. [†] (%)	Present (%)			Absent (%)		
				Good	Average	Bad	Good	Average	Bad
Positive effects									
D050 Psychological benefits	7.39**	2	6.4	60	5	35	35	31	34
D075 Humorous tone	6.09**	2	29.2	45	31	24	34	29	38
D125 No principal character	7.40**	2	12.2	32	47	21	38	27	36
D134 On-camera spokesperson	8.41*	4							
Voiceover only			55.8	32	32	36	46	25	29
Voiceover and characters			31.4	49	24	28	32	32	36
Characters only			7.1	32	32	36	38	29	33
D139 Framing of appeals	8.17*	4							
More negative			2.6	63	25	13	36	29	35
Balance positive/negative			7.7	58	21	21	35	30	35
More positive			89.7	34	30	36	59	22	19
Negative effects									
D080 Puffery	7.73**	2	5.4	24	12	65	38	30	32
D094 Demonstration in use	7.05**	2	76.6	36	26	38	40	38	22
D108 Music present	10.14**	2	85.6	33	30	36	58	22	20
D110 Music creates mood	8.30**	2	56.7	31	30	40	45	28	27

[†]Frequency is the proportion of ads in the data set that included the creative device.

** $p < 0.05$; * $p < 0.10$.

TABLE 2
Creative Devices Related to Sales and/or Persuasion Across Studies

Effects	Present Study	Stewart and Furse (1986)	Stewart and Koslow (1989)
Positive	D125 No principal characters D050 Psychological benefits D075 Humorous tone D134 Voiceover and characters D139 Negative appeals	D125 No principal characters	D024 News
		D024 News	D028 Double branded
		D028 Double branded	D079 Indirect comparisons
		D079 Indirect comparisons	D094 Demonstration in use
		D094 Demonstration in use	D095 Demonstration of results
		D095 Demonstration of results	D120 Principal character(s) actor
		D120 Principal character(s) actor	D140 Brand differentiating message
		D140 Brand differentiating message	
		D006 Sensory information	D019 Results information
		D022 Convenience information	D029 Manufacturer identified
		D032 Brand name reinforces use	D031 Auditory brand signoff
		D073 Relaxed/comfortable tone	D035 Setting directly related
		D089 Continuity of action	D042 Substantive supers
		X07 Time actual product is shown	D048 Ingredients as main message
Negative	None D080 Puffery D094 Demonstration in use D108 Music present D110 Music creates mood		D102 Problem solution
			D116 Principal character(s) female
		D013 Nutrition/health information	D013 Nutrition/health information
		D115 Principal character(s) male	D115 Principal character(s) male
		D127 Background cast	D127 Background cast
		X13 Number of character	X13 Number of character
		D008 Ingredients information	D020 User satisfaction or loyalty
		D136 Outdoor setting	D041 Surreal visuals
			D055 Enjoyment appeal
			D081 Front-end impact
			D098 Photographic stills
			D107 Visual pace
			X04 Time until category identified

Principal Components Analysis and Ordinal Regression

As with Stewart and Furse (1986), principal components analysis (PCA) with varimax rotation was used to reduce the number of creative devices and to remove the correlations between creative devices that are often used together in ads. Their analysis included approximately 74 creative devices and produced 24 creative factors that explained 62% of the variance among their ads' creative executions. Our analysis included 89 creative devices and produced 28 creative factors that explained 70% of the variance among our ads' creative executions. The factors and the primary loadings of the creative devices on these are presented in Table 3.

There are broad similarities in the factor structures between our data and Stewart and Furse's (1986). Factor scores were calculated for each ad for all of the 28 creative factors, including the entire matrix of factor score coefficients ($M = 0$, $SD = 1$). Seven ads were excluded because of missing data for one creative device, time until packaging or product is shown, because these ads did not show the package or product.

Relationship Between PCA Creative Factors and Sales Effectiveness

The ordinal regression was run with the factor scores as independent variables. Proportional odds were assessed by a likelihood ratio test comparing the fitted model to a model with varying location parameters, $\chi^2 = 28.59$, $p = .43$. The Pearson goodness-of-fit test indicated that the model was a reasonably good fit to the data, $\chi^2 (580) = 610.65$, $p = .18$. The model significantly predicted sales effectiveness over the intercept-only model, $\chi^2 (28) = 49.09$, $p < .01$, with a reasonable Nagelkerke R -square of .167.

To assess model fit we compared the model's sales effectiveness predictions to the observed data. The predictions were adequate, $\chi^2 (2) = 36.66$, $p < .001$, but only somewhat. The model correctly predicted the outcome (good, average, or bad) for only 50% of the data set. The model correctly predicted 73% of bad ads and 68% of good ads but could not predict mediocrity; all average ads were predicted as bad ads (44%) or good ads (56%). Because the model overdiscriminates the

TABLE 3
Factor Structure From PCA of Creative Devices

<u>Factor 1: Visual branding</u>	
Time until brand name is identified (%) (–.84)	Special offer (.70)
Time until product or pack is shown (%) (–.73)	Visual tagline (i.e., new information) (.67)
Times brand name or logo is shown (n) (.69)	How many product(s) advertised (.41)
Time until category is identified (%) (–.67)	<u>Factor 12: Negative appeals (with celebrities)</u>
Time brand name or logo is shown (%) (.60)	Principal character racial minority (.70)
Time package is shown (%) (.52)	Positive or negative appeals (–.70)
Blind lead-in (product identified at close) (–.43)	Principal character celebrity (.63)
<u>Factor 2: Humor (with characters)</u>	Psychological benefits, main message (.56)
Humorous tone (.78)	<u>Factor 13: Brand name</u>
Comedy or satire (.77)	Brand name reinforces product (.76)
Humorous closing (.71)	Double-branded product (.74)
Mood or image, dominant element (–.42)	Manufacturer is identified (.37)
On-camera spokesperson (.38)	<u>Factor 14: Continuity</u>
<u>Factor 3: Animated characters</u>	No principal character(s) (–.60)
Principal character animated (.88)	Continuity of action (.52)
Principal character brand created (.88)	Principal character(s) male (.42)
Animation or cartoon (all or most) (.69)	Length (.37)
Recognized continuing character (.69)	<u>Factor 15: Indoors</u>
<u>Factor 4: Ingredients</u>	Outdoor setting (–.91)
Ingredients as main message (.83)	Indoor setting (.87)
Information about ingredients (.82)	<u>Factor 16: Popular music (1.7%)</u>
Total propositions (.69)	Music as major element (.70)
Total information (.48)	Adaptation of well-known music (.69)
Time actual product is shown (%) (.40)	<u>Factor 17: Enjoyment (1.6%)</u>
Spoken words, per second (.36)	Total psychological appeals (.81)
<u>Factor 5: Product benefits</u>	Enjoyment appeals (.72)
Problem and solution format (.77)	Information about sensory experience (.37)
Information about results of using (.76)	<u>Factor 18: Fun reminder (1.6%)</u>
Demonstration of results of using (.64)	Happy/fun loving tone (.68)
Performance benefits, main message (.62)	Product reminder as main message (.43)
Use occasion (.51)	<u>Factor 19: Related setting (1.5%)</u>
<u>Factor 6: Emotions</u>	Setting unrelated but relevant (–.84)
Warm/caring tone (.73)	Setting directly related to product use (.56)
Cute/adorable tone (.72)	<u>Factor 20: Mood music (1.5%)</u>
Total emotional content (.68)	Music creates mood (.71)
Welfare appeals (.49)	Music present (.68)
Information about quality (.33)	<u>Factor 21: Cast (1.5%)</u>
<u>Factor 7: Fantasy</u>	How many characters (n) (.71)
Fantasy, exaggeration, or surrealism (.79)	Principal character(s) animal (–.50)
Surreal visuals (.78)	Background cast (.43)
Setting unrelated to product use (.58)	<u>Factor 22: Nutrition (1.3%)</u>
Demonstration of the product in use (–.39)	Wholesome/healthy tone (.72)
<u>Factor 8: Convenience</u>	Information about nutrition (.45)
Information about convenience (.77)	<u>Factor 23: Packaging (1.3%)</u>
Recognized, continuing musical theme (.75)	Information about packaging (.68)
Slice of life format (.51)	New/improved product or features (.47)
<u>Factor 9: Audio branding</u>	<u>Factor 24: Hard sell (1.3%)</u>
Auditory brand signoff (.75)	Hard sell tone (.76)
Memorable rhyme, slogan, or mnemonic (.65)	<u>Factor 25: Visual signoff (1.2%)</u>
Times brand name mentioned (n) (.54)	Dancing (–.67)
Pace, cuts per second (.37)	Visual brand signoff (.62)
<u>Factor 10: Female</u>	<u>Factor 26: Surprise (1.2%)</u>
Principal character female (.73)	Surprise middle (.63)
Principal character beautiful (.69)	Front-end impact (–.48)
Principal character actor playing role (.59)	<u>Factor 27: Supers (1.2%)</u>
<u>Factor 11: Ending information</u>	Substantive supers (i.e. text) (.79)
Auditory tagline (i.e., new information) (.80)	<u>Factor 28: Puffery (1.2%)</u>
	Puffery (.82)
	Research (.37)

TABLE 4
Ordinal Regression Results With PCA Factors Related to Sales

Factors	Exp <i>B</i>	Lower CI	Upper CI	<i>p</i>	Findings for Similar Factors From Stewart and Furse (1986) ^a
Positive effects					
Framing (Factor 12)	1.33	1.05	1.68	.02	Not comparable
Enjoyment (Factor 17)	1.22	.99	1.52	.07	Not comparable
Humor (Factor 2)	1.22	.98	1.52	.07	Factor 11: R/+, C/+
Animated characters (Factor 3)	1.22	.98	1.52	.08	Factor 6: C/–, P/–
Fantasy (Factor 7)	1.20	.97	1.50	.10	Factor 22: No effects
Visual brand signoff (Factor 25)	1.18	.95	1.47	.13	Factor 18: R/+
Related setting (Factor 19)	1.13	.91	1.41	.26	Factor 1: No effects
Indoor setting (Factor 15)	1.12	.91	1.39	.29	Factor 1: No effects
Popular music (Factor 16)	1.12	.90	1.38	.32	Factor 12: R/+, C/+
Hard sell (Factor 24)	1.10	.89	1.37	.37	Not comparable
Emotional tone (Factor 6)	1.08	.87	1.34	.48	Factor 10: No effects
Product benefits (Factor 5)	1.07	.87	1.33	.52	Factor 2: R/+, P/+
Convenience (Factor 8)	1.07	.86	1.34	.52	Factor 21: R/+, C/+, P/+
Ingredients (Factor 4)	1.06	.85	1.31	.61	Factor 4: R/–
Nutrition (Factor 22)	1.03	.83	1.28	.80	Factor 4: R/–
Negative effects					
Mood music (Factor 20)	.66	.53	.83	.00	Factor 12: R/+, C/+
Puffery (Factor 28)	.81	.65	1.01	.06	Not comparable
Ending information (Factor 11)	.83	.67	1.03	.09	Not comparable
Continuity (Factor 14)	.86	.69	1.07	.17	Factor 16: P/+
Packaging information (Factor 23)	.88	.71	1.10	.26	Not comparable
Fun reminder (Factor 18)	.89	.72	1.11	.30	Factor 20: P/–
Cast (Factor 21)	.91	.73	1.13	.40	Factor 5: C/–, P/–
Front-end impact (Factor 26)	.91	.74	1.13	.40	Not comparable
Audio branding (Factor 9)	.96	.78	1.20	.73	Factor 12: R/+, C/+
Female (Factor 10)	.96	.78	1.19	.73	Factor 15: R/–
Brand name (Factor 13)	.97	.78	1.20	.75	Factor 13: R/–, P/–
Supers (Factor 27)	.98	.79	1.22	.87	Factor 24: No effects
Branding prominence (Factor 1)	.99	.80	1.23	.92	Factor 3: C/+
Factor 7: R/+					

^aLetters used to denote the specific effectiveness measure(s) that the creative devices were related to, either positively or negatively: R = recall; C = comprehension; P = persuasion.

potential sales effectiveness of the average ads the findings should be approached with caution. The creative factors that affected model predictions are reported in Table 4, along with nonsignificant factors.

Negative appeals (Factor 12) had a strong positive relationship with sales effectiveness, whereas mood music (Factor 20) had a strong negative relationship. The negative appeals factor—which combined using some or primarily negative appeals, a racial minority, a celebrity (some celebrities were mixed race), and psychological benefits as the main message—increased the odds of sales effectiveness by 33%. The mood music factor (which combined the presence of music

and mood-creating music) decreased the odds of sales effectiveness by 34%.

Several other creative factors had marginally significant relationships with sales effectiveness. Enjoyment (combined enjoyment, a psychological appeal, with sensory information), humor (combined comedic elements with characters that speak), animated characters (combined animated and brand-created characters, which were present for several brands), and fantasy (combined unreal visuals, atypical settings, and no product demonstrations) all had modest positive effects. Puffery (combined unsubstantiated claims with research, e.g., 9 out of 10 dogs prefer Brand Z) and ending information

(combined the visual and/or audio presentation of new information in the closing seconds, e.g., retailers or other product varieties) each had modest negative effects.

Several of the results are not comparable across studies because there are creative devices present in our factors that were not analyzed by Stewart and Furse (1986) (e.g., message framing was insufficiently reliable) or vice versa. The result most consistent across studies, again, is the positive effect of humor. Two other results, however, appear to contradict Stewart and Furse (1986); We found that animated characters had a positive effect on sales, and music had a negative effect, whereas Stewart and Furse (1986) found the reverse on either recall, message comprehension, and/or persuasion. The animated factor from Stewart and Furse (1986), however, is not strictly comparable because it also captures the storyboard and animatic ads present in their data set, which suggests a more fundamental issue with the unfinished nature of the executions rather than a specific negative effect of animated/brand-created characters. The contradictory finding for music, however, is an interesting one. Stewart and Furse's (1986) factor also included slogans and mnemonics, which would explain how these might facilitate memory for the brand (recall), whereas our factor is more closely related to the prominence of music, which was too infrequent in Stewart and Furse's (1986) data to analyze.

DISCUSSION AND CONCLUSION

Many advertising researchers accept that creative or creativity matters. However, there is little generalized knowledge about sales-effective creative, with few studies conducted in real market settings on consumer behavior or sales. To help address this gap, we replicated Stewart and Furse's (1986) seminal codebook, which remains one of the most exhaustive deconstructions of advertising content, looking at the ability of different creative elements to distinguish between more and less sales-effective advertising.

Our replication showed that the codebook continues to work, with the relative frequencies of creative devices similar across data sets and time but with one stark difference: far fewer hard-sell tactics. This may be due to changing trends over time or differences in the respective data sets (e.g., product categories, countries) but regardless of the source, context and conditions are clearly important for this type of research. How often various creative devices are used by ads determines what can and cannot be investigated. The rarity of brand-differentiating messages in our data, for example, meant that we could not look at its relationship with sales. Variability across creative executions can make comparisons between studies difficult from the outset and some of the correlative findings may be idiosyncratic of the particular conditions captured by different data sets. However, if the findings hold over different data sets covering different conditions then this concern fades.

Our analysis of the sales impact of separate creative devices and the more elaborate creative factors showed some similarities to but many differences from Stewart and colleagues' studies. Humor was the strongest directionally consistent result between the original studies and ours for improving advertising effectiveness, though not on directly comparable measures. Humor was also more common in our data set, 29% versus 5%, which suggests a more pervasive effect across ads. The standout difference was the negative effect of music on sales. Perhaps prominent mood-creating music can be distracting for the viewer, preventing some level of audio-message processing and/or precluding other important audio elements; for example, music reduces the opportunity for speaking characters, which had a positive effect on sales. Analyzing the different interactions of creative devices, which was not possible here due to sample size limitations, may reveal synergistic or counteracting relationships. Investigating combinations of creative devices linked to sales effects presents an interesting and challenging avenue for future research.

The low congruence between our findings and those of Stewart and colleagues is perhaps to be expected given the differences in the dependent variables. The ARS persuasion measure is related to sales, but it is not a perfect correlation (Blair and Kuse 2004). Notably the two original studies did not themselves produce highly congruent results (see Table 2) despite being near perfect replications (same data source and country, similar categories).

Perhaps, then, the most important finding now substantiated in terms of sales effectiveness is that "[t]here is no magic formula for the creation of effective advertising. While some general guidelines for creating effective advertising appear to receive support from the findings, it is also true that no one executional factor accounts for much of the total variance of any of the measures of advertising effectiveness" (Stewart and Koslow 1989, p. 30). This will be an obvious finding to some and surprising to others, just as some people think it is relatively easy to spot good (effective) or bad (ineffective) advertising, whereas others view advertising as a mysterious dark art. The evidence, however, supports the middle ground. Our findings do provide support for some of the results of Stewart and colleagues, and those of others, regarding which creative devices matter the most. So even though they are correlational findings, collectively they strongly suggest the possibility of causality that should be further investigated in laboratory and field studies.

Armstrong (2010) argues that nonexperimental studies such as those we have replicated, even when done on a large-scale and by competent researchers, are unlikely to advance knowledge in this area because of such small, correlational effects. Armstrong (2010), however, still accepts findings from Stewart and colleagues' studies as supporting evidence for 20 of his 171 persuasion principles for motion media—more than most other studies that are cited. We suggest large-scale studies of the performance of real ads measured in market do have more

relevant insights to offer about creative effects, especially as we move into a “big data” era. Very large data sets of many different creatives with behavioral response measures will eventually become more commonplace. This will provide researchers with new opportunities to hold at least some variables constant (quasi-experiments) to progressively examine the efficacy of creative devices in diverse contexts among different consumer groups outside of the laboratory. To put this issue into perspective, however: Though small, the in-market effects of creative devices are worth investigating considering that advertising creative decisions are one of the few things that marketers can control to deliver incremental sales above a mostly stable baseline, which is largely an outcome of history, inertia, and existing customer loyalties.

Our findings further suggest creative devices that are related to sales warrant more research. The coding here was quite general; as such, coding with more specificity related to the various types of humor (e.g., silly, sarcastic, puns) or types of music (e.g., genres, but also structural elements, such as tempo) should help us better understand these effects on sales. There is already a quite substantial body of research dedicated to music and humor in advertising, but most studies have looked exclusively at intermediate effects rather than brand choice or sales (see Allan 2014; Eisend 2009).

Though not a magic formula, our findings offer further empirical support for guidelines that advertisers can consider when developing new creative. Most broadly, many of our findings speak to the importance of creative execution devices rather than message strategy devices. *How* advertisers chose to communicate (e.g., framing, characters, situations) is possibly more impactful than *what* is communicated (e.g., news, features, or benefits). The importance of execution devices is consistent with the theory of transformational advertising for low-involvement products (Rossiter and Bellman 2005), which our data represent. Our data specifically suggest that advertisers should look to emphasize what can happen when the product is *not* used and humorous elements. Also, advertisers should carefully consider their selection of music, ideally earlier in the creative development process, which is reportedly far from the norm (Binet, Müllensiefen, and Edwards 2013). Any creative work that runs counter to such empirically derived guidelines should be subjected to some form of evaluative testing.

FUNDING

We would like to acknowledge Mars Incorporated for providing the data and funding, as well as for their advice and support throughout this project.

SUPPLEMENTAL DATA

Supplemental data for this article can be accessed at www.tandfonline.com/ujoa.

REFERENCES

- Allan, David (2014), “Turn It Up: That’s My Song in That Ad,” *International Journal of Music Business Research*, 3 (1), 26–51.
- Armstrong, J.S. (2010), *Persuasive advertising: Evidence-based principles*, Basingstoke: Palgrave Macmillan.
- Bellman, Steven, Anika Schweda, and Duane Varan (2012), “Interactive TV Advertising: iTV Ad Executional Factors,” *Journal of Business Research*, 65 (6), 831–39.
- Binet, Les, Daniel Müllensiefen, and Paul Edwards (2013), “The Power of Music,” *Admap*, October, www.warc.com/admap.
- Blair, Margaret H., and Allan R. Kuse (2004), “Better Practices in Advertising Can Change a Cost of Doing Business to Wise Investments in the Business,” *Journal of Advertising Research*, 44 (1), 71–89.
- Chandy, Rejesh K., Gerard J. Tellis, Deborah J. MacInnis, and Pattana Thaivanich (2001), “What to Say When: Advertising Appeals in Evolving Markets,” *Journal of Marketing Research*, 38 (4), 399–414.
- Eisend, Martin (2009), “A Meta-Analysis of Humor in Advertising,” *Journal of the Academy of Marketing Science*, 37 (2), 191–203.
- , Julia Plagemann, and Julia Sollwedel (2014), “Gender Roles and Humor in Advertising: The Occurrence of Stereotyping in Humorous and Nonhumorous Advertising and Its Consequences for Advertising Effectiveness,” *Journal of Advertising*, 43 (3), 256–73.
- Haley, Russell, and Allan Baldinger (1991), “The ARF Copy Research Validity Project,” *Journal of Advertising Research*, 31 (2), 11–32.
- Heath, Robert (2009), “Emotional Engagement: How Television Builds Big Brands at Low Attention,” *Journal of Advertising Research*, 49 (1), 62–73.
- Holbrook, Morris, and Donald Lehmann (1980), “Form versus Content in Predicting Starch Scores,” *Journal of Advertising Research*, 20 (4), 53–62.
- Jones, John Philip (1995), *When Ads Work: New Proof That Advertising Triggers Sales*, 1st ed., New York: Lexington Books.
- Lehnert, Kevin, Brian D. Till, and José M. Ospina (2014), “Advertising Creativity: The Role of Divergence versus Meaningfulness,” *Journal of Advertising*, 43 (3), 274–85.
- Lodish, Leonard M., Magid Abraham, Stuart Kalmenson, Jeanne Live-Isberger, Beth Lubetkin, Bruce Richardson, and Mary E. Stevens (1995), “How TV Advertising Works: A Meta-Analysis of 389 Real World Split Cable TV Advertising Experiments,” *Journal of Marketing Research*, 32 (2), 125–39.
- McEwen, William J., and Clark Leavitt (1976), “A Way to Describe TV Commercials,” *Journal of Advertising Research*, 16 (6), 35–39.
- McQuarrie, Edward, Barbara Phillips, and Steven Andrews (2012), “How Relevant Is Marketing Scholarship? A Case History With a Prediction,” *Advances in Consumer Research*, 40, 342–49.
- Phillips, Diane M., and John L. Stanton (2004), “Age-Related Differences in Advertising: Recall and Persuasion,” *Journal of Targeting, Measurement, and Analysis for Marketing*, 13 (1), 7–20.
- Rossiter, John R., and Steven Bellman (2005), *Marketing Communications: Theory and Applications*, Frenchs Forest, Australia: Pearson Education.
- Rust, Roland T., and Bruce Cooil (1994), “Reliability Measures for Qualitative Data: Theory and Implications,” *Journal of Marketing Research*, 31 (1), 1–14.
- Sasser, Sheila L., and Scott Koslow (2008), “Desperately Seeking Advertising Creativity: Engaging an Imaginative ‘3Ps’ Research Agenda,” *Journal of Advertising*, 37 (4), 5–20.
- Smith, Robert E., and Xiaojing Yang (2004), “Toward a General Theory of Creativity in Advertising: Examining the Role of Divergence,” *Marketing Theory*, 4 (1–2), 31–58.
- Stanton, John L., and Jeffrey Burke (1998), “Comparative Effectiveness of Executional Elements in TV Advertising 15 versus 30 Second Commercials,” *Journal of Advertising Research*, 38 (6), 7–13.
- Stewart, David W. (2009), “Marketing Accountability: Linking Marketing Actions to Financial Results,” *Journal of Business Research*, 62 (6), 636–43.

- , and David H. Furse (1984), "Analysis of the Impact of Executional Factors on Advertising Performance," *Journal of Advertising Research*, 24 (6), 23–26.
- , and ——— (1985), "The Effects of Television Advertising Execution on Recall, Comprehension, and Persuasion," *Psychology and Marketing*, 2 (3), 135–60.
- , and ——— (1986), *Effective Television Advertising: A Study of 1000 Commercials*, Lexington, MA: Lexington Books.
- , and Scott Koslow (1989), "Executional Factors and Advertising Effectiveness: A Replication," *Journal of Advertising*, 18 (3), 21–32.
- Tellis, Gerard J. (2009), "Generalizations About Advertising Effectiveness in Markets," *Journal of Advertising Research*, 49 (2), 240–45.
- Theodorakis, Ioannis G., Christos Koritos, and Vlasios Stathakopoulos (2014), "Rhetorical Maneuvers in a Controversial Tide: Assessing the Boundaries of Advertising Rhetoric," *Journal of Advertising*, 44 (1), 14–24.
- Twedt, Dik W. (1952), "A Multiple Factor Analysis of Advertising Readership," *Journal of Applied Psychology*, 36 (3), 207–15.
- Vakratsas, Demetrios, and Tim Ambler (1999), "How Advertising Works: What Do We Really Know?," *Journal of Marketing*, 63 (1), 26–43.
- Wood, Leslie (2009), "Short-Term Effects of Advertising: Some Well-Established Law-Like Patterns," *Journal of Advertising Research*, 49 (2), 186–92.
- Yoon, Hye Jin, and Yeuseung Kim (2014), "The Moderating Role of Gender Identity in Responses to Comedic Violence Advertising," *Journal of Advertising*, 43 (4), 382–96.