

copy test methods to pretest advertisements

Cornelia Pechmann and J. Craig Andrews

Q1

In this paper, we define advertising copy testing, describe the major copy test methods using case examples, discuss each method's strengths and limitations, and identify areas of debate. The case examples were selected based on the authors' direct experiences and the methodological rigor of the methods. We define advertising copy testing as a quantitative and qualitative marketing research method that is used to pretest advertisements where numerical data on effects is collected and statistically analyzed. The key industry document about it called PACT (1982) states: "Copy testing is undertaken when a decision is to be made about whether the advertising should be run in the marketplace. Whether this stage utilizes a single test or a combination of tests, its purpose is to aid in the judgment of specific advertising executions" (page 8).

Generally, there are four stages in advertising research: (i) copy development, (ii) a rough stage, (iii) copy testing (i.e., pretesting), and (iv) tracking (i.e., posttesting) (Belch and Belch, 2007; Shimp, 2010). In copy development, initial ideas are developed from the creative brief and screened in focus groups. In the rough stage, the advertisement is tested in draft or storyboard form, again in focus groups. Next, the final or nearly final advertisements are copy tested. Finally, after appearing in the media, the advertisements are tracked using surveys and/or sales data. The surveys typically assess advertising recall and recognition, although some measure advertising attitudes, product beliefs, and/or purchase intent.

Nine principles of copy testing were developed by US advertising agencies and are set forth in PACT (1982, p. 10–27). PACT states that for a copy test method to be sound, it must (i) be relevant to the objectives, (ii) have agreement about the use of results, (iii) use multiple measurements, (iv) be based on a model of human response, (v) consider multiple exposures, (vi) test comparably finished executions, (vii) control the exposure context, (viii) define the relevant sample, and (ix) demonstrate reliability and validity. To this list, we add (x) take baseline (i.e., pre-exposure) measurements

and/or use control groups. We add this because it is in accordance with sound and generally accepted research standards (Cook and Campbell, 1979). A copy test may use a prepost design that compares participants' responses before and after they see the advertisement. Or, a copy test may randomly assign participants to a test group that sees the advertisement, or to a control group that does not, and compare their responses. Some designs make use of both approaches (Malholtra, 2009). We now describe three major copy test methods that apply these principles that are used by the US National Youth Anti-Drug Media Campaign, the US Federal Trade Commission, 1983, and the ARSgroup, a major copy test firm.

COPY TESTING FOR THE NATIONAL YOUTH ANTI-DRUG MEDIA CAMPAIGN

Q2

METHOD

The US Office of National Drug Control Policy, with the assistance of The Partnership for a Drug-Free America, has run the National Youth Anti-Drug Media Campaign for about a decade. Total expenditures are over a billion dollars (Foley and Pechmann, 2004). The antidrug advertisements target either the youth or their parents. Each advertisement is copy tested prior to airing using a method that was developed by academic experts who conduct similar copy tests for their own research (e.g., Pechmann and Reibling, 2006).

For each ad tested in a youth copy test, the sample consists of 200 Caucasians, 200 Hispanics, and 200 African-Americans, split 50/50 on grade (grades 7–8 vs. grades 9–10) and 50/50 on gender. From each ethnic group and age, 25 youths are randomly assigned to the test group (they see the antidrug advertisement) and 25 are randomly assigned to the nonexposure control group (they do not see the advertisement). For each ad tested in a parent copy test, the sample consists of 100 Caucasians, 100 Hispanics, and 100 African-Americans, with at least one child in grades 7–10, and split about 50/50 on gender. From each ethnic group, 50 parents are randomly assigned to the test group and 50 are randomly assigned to the control group. The sample sizes were chosen to detect

2 copy test methods to pretest advertisements

differences of 10% or more between test and control groups with a power of 0.80, a 95% confidence level ($p < 0.05$), and an observed within-group variance of 0.6.

The data collection is subcontracted to an experienced marketing research firm and is typically completed in one weekend using about 40 shopping malls across the United States where the firm has access to research facilities. Youth are prescreened based on grade, ethnicity, and gender and their sensation seeking is also assessed. Parents are prescreened based on their children's grades and their own ethnicity. Those who pass the screen and give informed consent are paid \$1 for completing the 10-minute study and the response rate is about 78%. Parental consent for youths' participation is obtained whenever feasible. The study is conducted in small rooms with trained English-speaking interviewers who read the questions to individual participants and record their answers. Typically multi-item scales with 5 to 10 point response options are used.

Test participants see the antidrug advertisement two times in succession, while control participants do not see it. Next, test participants answer open-ended questions about what the advertisement is about and their responses are later coded. Test participants also rate the advertisement on diagnostic measures of effectiveness, believability, attention-getting power, and argument strength. The advertisement's ratings are compared with norms from prior testing. Then, all participants complete a questionnaire that includes measures of their drug-related beliefs and intent. The measures are pretested to ensure clarity and sensitivity (no ceiling/floor effects) and are reused across multiple copy tests to ensure comparability. The data are analyzed using analysis of covariance; there is one independent variable (test vs. control condition), moderators or contingent factors (for youth: grade, gender, and ethnicity; for parents: ethnicity) and several covariates (for youth: exact age and sensation seeking score; for parents: the youth's grade). The effects of test versus control condition alone and in combination with grade, gender, and/or ethnicity are examined by looking at main effects and two- and three-way interactions. This analysis is done for each belief and intent measure.

A youth advertisement is recommended for airing if it significantly ($p < 0.05$) strengthens an antidrug belief or weakens intent to use drugs among the overall sample or among a subsample based on grade, gender, or ethnicity. It is not recommended for airing if it weakens an antidrug belief or strengthens intent to use drugs. A parent advertisement is recommended for airing if it significantly ($p < 0.05$) strengthens an antidrug parenting belief or intent among the overall sample or a subsample based on gender or ethnicity. It is not recommended for airing if it weakens an antidrug parenting belief or intent. The diagnostic measures help reveal an advertisement's specific strengths or weaknesses.

STRENGTHS

The National Youth Anti-Drug Media Campaign's copy test method seems to comply with all 10 principles of good copy testing that were outlined above.

1. *Be relevant to the objectives:* The campaign objectives are to strengthen antidrug beliefs and intent and avoid adverse effects, and the copy testing explicitly tests this.
2. *Have agreement about the use of results:* There are clear criteria for recommending or not recommending advertisements for airing.
3. *Uses multiple measurements:* There are multiple measures of the ad, beliefs, and intent.
4. *Be based on a model of human response:* Well-established models are used, such as hierarchy of effects, the elaboration likelihood model, attitude toward the ad, and steps in information processing (Vakratsas and Ambler, 1999). For example, beliefs are predicted to affect intent.
5. *Consider multiple exposures:* Participants see the antidrug advertisement twice.
6. *Test comparably finished executions:* Virtually all advertisements are tested in final form.
7. *Control the exposure context:* The antidrug advertisement is shown alone, without clutter.
8. *Define the relevant sample:* The target grades, genders, and ethnicities are stated.

copy test methods to pretest advertisements 3

9. *Demonstrate reliability and validity:* Multi-item measures are used to assess reliability. Also, copy test results are compared with tracking results to assess validity. Both reliability and validity have been found to be high.
10. *Take baseline (i.e., pre-exposure) measurements and/or use control groups:* A randomly selected control group does not see the antidrug advertisement, but just completes the belief and intent measures, and statistical comparisons are made between control and test groups. Compared with pretest designs, control groups help rule out alternative explanations for reported antidrug beliefs and intent, such as mere measurement effects and cueing of other outside campaigns simply caused by asking about antidrug beliefs and intent. In addition, a pretest design was ruled out because pretest responses are likely to bias posttest responses; the measurements are too close in time (Cook and Campbell, 1979). Specific biases of concern are that people may purposely report the same beliefs at both points in time or exhibit reactance against the testing and claim adverse effects.

Q3

LIMITATIONS

In the National Youth Anti-Drug Media Campaign's copy tests, youths' interest in drug use is not measured. Sensation seeking, which is related to drug use, is measured based on youth activities and interests (e.g., "do scary things," "explore strange places," "break rules") (Palmgreen *et al.*, 2001). Sensation seeking is included as a covariate to ensure that test and control groups are equated on it. Drug use could be directly measured so that the advertising effects could be assessed for drug user and nonuser subsamples, but this would add to the cost and sample size needs. Sample sizes are already very small for three-way interactions with just 50 participants per cell.

There is no correction for conducting multiple comparisons across the different belief and intent measures. The significance level is set at the standard $p < 0.05$, which means that on average 1 in 20 measures will show a significant effect due

to chance alone, although this could be avoided by using a simple Bonferroni test (Winer *et al.*, 1991). Also, control participants do not see a control advertisement; they see nothing. It is generally recommended that control participants be given a pseudotreatment or placebo to avoid the Hawthorne effect in which being a full research participant itself improves responses. Specifically, a control advertisement such as a nondrug public service announcement could be used and control participants could answer questions about it (Andrews and Maronick, 1995, Pechmann and Reibling, 2006). However, the control advertisement must be truly neutral. Another issue is that some involved in the campaign think that an antidrug advertisement should be aired if it works on the target group, even if it may have adverse effects on a nontarget group. However, this seems problematic because often media spills over from one group to another. Finally, some involved in the campaign think advertisements that test poorly can be fixed and aired without retesting, but retesting it helps to ensure that problems are truly resolved.

METHOD

The US Federal Trade Commission (FTC) has employed copy testing in advertising deception cases for over 35 years (Maronick, 1991). Although it is not required per se, extrinsic evidence such as copy testing is usually considered at the FTC and given substantial weight in cases involving potentially deceptive implied ad claims (Andrews and Maronick, 1995). Extrinsic evidence can come in the form of methodologically sound consumer research studies, such as copy tests (Kraft, Inc., 1991), marketing research, and/or expert testimony.

Generally accepted copy test principles exist at the FTC, as well as substantial case precedent for FTC copy tests (e.g., Kraft, Inc., 1991, Stouffer Foods Corp., 1994; Thompson Medical Company Inc., 1984), but how these principles are operationalized may vary somewhat in each specific case (Maronick, 1991). Such generally accepted principles for copy testing should include the following: (i) the proper universe is examined, (ii) a representative sample is drawn from the universe, (iii) the mode of questioning is correct, (iv)

4 copy test methods to pretest advertisements

the survey is designed by recognized experts, (v) the data gathered are accurately reported, and (vi) the sampling design, questionnaire, interviewing, and statistical analyses are in accordance with general accepted standards for conducting survey research (Manual for Complex Litigation, 1982).

Usually, an FTC advertising deception case begins with a complaint involving interstate commerce in the public interest (Andrews, 2001). Then, discovery is undertaken to obtain ads, media schedules, marketing research, and marketing plans; and an in-house copy test might be conducted. At a later point, the FTC might conduct an outside copy test with external marketing consultants if it appears that the case might go to trial. Typically, the copy test questions are pretested, followed by a main study with about 100 participants in the test group who see the challenged advertisement, and an additional 100 in the control group who see a control advertisement instead. Participants in the control group can be randomly assigned to see (i) a purged advertisement that is identical to the challenged advertisement, but with the challenged claims removed, (ii) a different advertisement for the same brand that does not contain the challenged claims (Kraft, Inc., 1991), or (iii) no advertisement, that is, a no-exposure control (Andrews and Maronick, 1995). Data collection is usually subcontracted to an experienced firm that uses research facilities in at least three geographically dispersed shopping malls.

A standard FTC copy test begins with screener questions to ensure that the sample is representative of the universe. In the case of Stouffer Foods Corp. (1994), the sample was drawn from a universe of principal food shoppers between the ages of 25 and 54 who had purchased a frozen entrée in the last three months and were not on a medically supervised diet. Once participants are successfully screened, they are shown the challenged advertisement or a control advertisement; it is shown with two clutter advertisements. Then, participants answer the survey questions, which generally follow a “funnel approach,” beginning with general open-ended questions and then moving to successively narrower questions and ending with specific closed-ended ones (Kraft, Inc., 1991).

It is important that the questions be related to the advertising claims; they cannot just measure general beliefs about a product attribute.

Hypothetically, an FTC case might be about a potentially misleading, low-fat advertising claim. After the respondents have finished looking at the target advertisement with clutter ads, a first set of questions would ask whether they remember seeing the advertisement and the brand in question. If answered correctly, they would be asked, “What does the brand’s ad say or suggest to you about its products?” This would be followed with probes of “anything else” and would also include a “don’t know” option. Next, respondents would be asked to read the target advertisement one more time. When finished, they would be queried, “Does the ad say or suggest anything about the amount of fat in the brand?” with the response options of yes, no, and, don’t know/maybe/not sure. If the answer is “yes,” they would be asked, “What does the ad say or suggest to you about the amount of fat in the brand’s products?” with probes and don’t know options. Respondents would then be asked, “Does the ad say or suggest that the brand’s products are . . . ” with the options presented as “low-fat foods,” “high-fat foods,” or “neither low- nor high-fat foods,” as well as a “don’t know” option.

Toward the end, other specific closed-ended and/or control questions may be presented. For instance, a control question is often asked involving an attribute that is not related to the challenged advertisement or the control advertisement either (e.g., about sugar in the case of challenged low-sodium claims, see Stouffer Foods Inc., 1994). The purpose of this is to help account for “yea-saying” on behalf of respondents. Lastly, demographic questions are collected. In the data analyses, the response percentages are computed, and then a net response is calculated by subtracting the responses to the control advertisement from the responses to the test advertisement. Sometimes, the respondent verbatims from open ends are summarized and presented along with the statistical results. Copy testing for corrective advertising is more involved. In the *FTC v. Novartis* case on Doan’s Pills (Mazis, 2001), the tests assessed deception, claim materiality, false

copy test methods to pretest advertisements 5

beliefs, ad-belief linkages, lingering beliefs, and remedies.

STRENGTHS

The FTC's copy test method seems to comply with all 10 principles of good copy testing.

1. *Be relevant to the objectives:* The copy testing assesses the consumer's reasonable interpretation and net impression of potentially misleading ad claims that are material or important (i.e., likely to affect conduct).
2. *Have agreement about the use of results:* The copy testing decision rules are consistent with generally acceptable principles, case precedent, and the 1983 FTC Deception Policy.
3. *Use multiple measurements:* A funneling process is used, from general open-ended questions to increasingly narrower questions and ending with specific closed-ended questions.
4. *Be based on a model of human response:* Copy testing is based on an information-processing model that recognizes the steps of exposure, awareness, comprehension, persuasion, and intent, though the focus is on comprehension (i.e., beliefs).
5. *Consider multiple exposures:* Participants see the test or control advertisement twice, once before general open-ended questions and a second time before more specific questions.
6. *Test comparably finished executions:* Advertisements are copy tested in final form, as they appeared in the media.
7. *Control the exposure context:* The test or control advertisement is shown with two clutter advertisements during the first exposure and alone during the second exposure.
8. *Define the relevant sample:* Screener questions are used to ensure that the sample is in the target market and demographics are assessed as well.
9. *Demonstrate reliability and validity:* Although single-item measures are commonly used, the funneling process allows for validity checks across measures. Also, control questions and don't know options help minimize the yea-saying bias.

10. *Take baseline (i.e., pre-exposure) measurements and/or use control groups:* Some participants are randomly assigned to see a control advertisement that does not contain the challenged claims.

LIMITATIONS

The FTC copy testing also involves tradeoffs. It can be difficult to examine low salience or tertiary claims if awareness of these claims is low in both the open- and closed-ended questions (Maronick, 1991). Also, on occasion, there could be strong halo effects; anything in the advertisement might be considered part of the challenged claims or might immediately elicit memories of a longer running campaign. This situation makes it difficult to create a control advertisement that excises or excludes the challenged claims (Andrews and Maronick, 1995). Researchers could use a no-ad-exposure control group, but the standard FTC questions are ad specific. Also, the FTC typically uses single item versus multi-item measures, which precludes standard reliability checks (i.e., alpha). However, the funneling procedure allows for some validity assessment across measures. Finally, most data is presented in the form of simple percentage comparisons between test and control groups or between test and control questions. Thus, there could be room for more sophisticated interval-scaled measures and analyses.

METHOD

The ARSgroup of The Quality Measurement Company has copy tested television advertisements for over 40 years. ARSgroup has conducted over 40 000 copy tests worldwide, often for major firms such as Campbell Soup, Procter and Gamble, and Wal-Mart. It specializes in gauging the effectiveness of television advertisements based on its ARS persuasion score. Its competitors include Nielsen IAG, The Pretesting Company, and Marketing Evaluations, among others. The information below is based on published articles about ARSgroup's advertising copy testing method (Adams and Blair, 1992; Blair and Rabuck, 1998; Stewart *et al.*, 1983). Overall, the ARSgroup's

6 copy test methods to pretest advertisements

copy test research shows that advertising quality, as measured by their ARS persuasion score, is related to short-term (i.e., weekly or monthly) changes in brand sales and share. Also, advertising quality is more important than advertising expenditures in affecting sales and share. Finally, a high-quality advertisement builds sales quickly, but has diminishing returns over time because its effect gradually wears out.

Typically, the ARSgroup copy tests a television advertisement in four geographically dispersed cities (Adams and Blair, 1992; Stewart *et al.*, 1983). Approximately 400–1000 randomly selected men and women are recruited to participate ostensibly to evaluate a television pilot and are instructed to report to a research facility on a specific date and time. As an incentive, participants are told that raffles will be conducted and that many of them will win free bags of products. When they arrive, participants are shown lists of brands and are asked to choose the brand they would most like to win in each product category. Next, a raffle is conducted and some people win the products they have chosen. Then, participants view a television program in which each test advertisement is embedded along with unrelated or noncompeting clutter advertisements. At the end, participants are shown the same lists of brands again, and are asked to choose the brands that they would most like to win this time. Again, a raffle is conducted and some people win the products they have chosen. For this reason, virtually all ARSgroup copy tests involve low-cost consumer packaged goods.

The raw ARS persuasion score is calculated as the percentage of participants who choose the test brand after exposure to the test advertisement, minus the percent who chose the test brand before exposure to the test advertisement. This raw score is then converted to an adjusted persuasion score; the ARSgroup has found that the brand's market share biases the copy test responses and so this bias is removed. Three days after the copy test, participants are interviewed by phone to assess advertising recall (Stewart *et al.*, 1983), but the main measure used to evaluate advertisements is the adjusted ARS persuasion score. The ARSgroup's reliability studies indicate the adjusted persuasion score's test-retest reliability is 0.93 (Stewart

et al., 1983). Its validity studies show the adjusted persuasion score predicts trial rates for new, frequently purchased, branded consumer goods ($r = 0.85$). ARSgroup has also validated its copy test method for established brands in split cable ad tests, meaning some households receive the advertisement while others do not, and household purchase rates are compared (Adams and Blair, 1992).

STRENGTHS

The ARSgroup method generally complies with the 10 principles of good copy testing.

1. *Be relevant to the objectives:* The ARSgroup objective is to predict an advertisement's effect on brand sales and share; its validation studies indicate that it does this.
2. *Have agreement about the use of results:* ARSgroup's recommendations are clear – ads with low ARS persuasion scores should not be used because they will not improve sales no matter how much they air.
3. *Use multiple measurements:* The ARSgroup measures product choice as well as advertising recall.
4. *Be based on a model of human response:* There is no explicit model of advertising response.
5. *Consider multiple exposures:* Participants see an ad just once, but the ARSgroup's validation studies suggest this is sufficient (Adams and Blair, 1992; Blair and Rabuck, 1998).
6. *Test comparably finished executions:* The ARSgroup copy tests television ads in final or near final form; it does not test storyboards.
7. *Controls the exposure context:* The exposure context is carefully controlled. The advertisements are embedded in the same television program with the same clutter advertisements across multiple copy tests.
8. *Define the relevant sample:* A general audience sample is used, but most of the copy tests are for general or mass-market products.
9. *Demonstrate reliability and validity:* The ARSgroup has demonstrated reliability and validity.
10. *Take baseline (pre-exposure) measurements and/or use control groups:* The ARSgroup

copy test methods to pretest advertisements 7

takes a pre-exposure measure of brand choice and uses this as the baseline for assessing the advertisement's persuasive impact on brand choice.

LIMITATIONS

The ARSgroup does not disclose the adjustments they make to the raw persuasion score to correct for response biases because of brand market share; their adjustments are proprietary. This makes it difficult to fully evaluate the method and to publish ARSgroup studies; most academic journals require that measurement adjustments be disclosed. This proprietary approach also prohibits independent replication of the ARSgroup validation studies. In addition, many copy test participants receive the products they choose and so the advertised products must be low-cost consumer packaged goods. The ARSgroup method cannot be used for expensive items such as cars, or for illegal items such as marijuana. Since a representative sample is employed, the products must have reasonably wide appeal.

There are many general areas of debate about copy testing. The main ones are (i) the reliance on convenience samples such as people in shopping malls, (ii) the artificial viewing environment including forced exposure to the advertising, (iii) no assessment of cumulative advertising effects, (iv) often no measure of actual behavior, (v) no assessment of advertising-induced word of mouth, and (vi) the increased use of on-line or web-based copy testing, which means less control over samples and exposure environments, although the effective use of screeners can help. A longstanding concern is the reliance on malls and other public venues that offer convenience samples. However, convenience sampling is common in marketing research and the US courts have accepted it (Maronick, 1991; Plevan and Siroky, 1991). The artificial viewing environment is of concern and so many copy tests include clutter. The FTC shows the test advertisement with two clutter advertisements and then measures awareness of it. The ARSgroup embeds the test advertisements with numerous clutter advertisements.

Copy tests typically offer no assessment of cumulative advertising effects, but most ensure

that consumers have seen the test advertisement at least twice. Research suggests that an advertisement generally has its impact after just two or three exposures (Pechmann and Stewart, 1988). The ad's effect on actual product use or purchase behavior should be measured whenever possible, if this is relevant to the copy test objectives. The FTC copy tests often aim to measure false advertising beliefs. For other copy tests, behavior can often be measured if the budgets are adequate and the test advertisements promote low-cost consumer goods. For expensive items such as cars, illegal items such as marijuana, and high-risk items such as cigarettes, it can be very difficult to assess behavior. In such cases, it is important to use validated intent measures.

Copy tests generally do not measure word of mouth effects, for instance, when one person talks to another about a memorable advertisement. Word of mouth may produce either positive or negative outcomes. The ARSgroup's copy tests involve large groups of people who could be permitted to talk about the ads, but this would likely be too contrived. Internet postings may be a better way to assess word-of-mouth advertising effects. Also, a recent concern is that on-line or web-based copy testing might be less valid than traditional mall-based methods. People who participate in on-line copy tests may be less representative than people in malls, although effective screeners can help in this regard. Yet, the on-line environment can be less controllable than a mall facility; for instance, people on-line can talk to others and/or multitask while completing the survey. More research is needed to compare on-line versus mall-based copy tests.

In summary, copy testing is a standard and well-accepted marketing research method that has been shown to be reliable and valid if it is done well. Companies and other entities should feel confident about using copy testing to evaluate advertisements as long as the recommended procedures discussed above are followed.

Bibliography

- Adams, A.J. and Blair, M.H. (1992) Persuasive advertising and sales accountability: past experience and forward validation. *Journal of Advertising Research*, 32 (2), 20–25.
- Andrews, J.C. (2001) The use of marketing knowledge in formulating and enforcing consumer protection

8 copy test methods to pretest advertisements

- policy, in *Handbook of Marketing and Society* (eds P.N. Bloom and G.T. Gundlach), Sage Publications, Thousand Oaks, CA, pp. 1–33.
- Andrews, J.C. and Maronick, T.J. (1995) Advertising research issues from FTC versus stouffer foods corporation. *Journal of Public Policy and Marketing*, **14**, 301–327.
- Belch, G.E. and Belch, M.A. (2007) *Advertising and Promotion: An Integrated Marketing Communications Perspective*, 7th edn., McGraw-Hill Irwin, Boston, MA.
- Blair, M.H. and Rabuck, M.J. (1998) Advertising wearin and wearout: ten years later—more empirical evidence and successful practice. *Journal of Advertising Research*, **38** (5), 7–18.
- Clark Boardman Company Ltd (1982) *Manual for Complex Litigation*, Federal Practices Series, Clark Boardman Company Ltd., New York, NY.
- Cook, T.D. and Campbell, D.T. (1979) *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, Houghton Mifflin, Boston, MA.
- U.S. Federal Trade Commission (1983) Policy Statement on Deception October 14, 1–19; appended to *Cliffdale Associates, Inc.* (1984), Commission Decision, 103 FTC 174–184.
- Foley, D. and Pechmann, C. (2004) The national youth anti-drug media campaign copy test system. *Social Marketing Quarterly*, **10**, 34–42.
- Kraft, Inc. (1991) Commission decision, 114 FTC 40–46, pp. 116–151.
- Maholtra, N.K. (2009) *Basic Marketing Research: A Decision-Making Approach*, Prentice Hall, Upper Saddle River, NJ.
- Maronick, T.J. (1991) Copy tests in FTC deception cases: guidelines for researchers. *Journal of Advertising Research*, **31**, 9–17.
- Mazis, M.B. (2001) FTC v. Novartis: the return of corrective advertising? *Journal of Public Policy & Marketing*, **20**, 114–122.
- PACT (1982) Positioning advertising copy testing. *Journal of Advertising*, **11** (4), 4–29.
- Palmgreen, P., Donohew, L., Lorch, E. *et al.* (2001) Television campaigns and adolescent marijuana use: tests of sensation seeking targeting. *American Journal of Public Health*, **91** (2), 292–296.
- Pechmann, C. and Reibling, E. (2006) Antismoking advertisements for youth: an independent evaluation of health, counter-industry, and industry approaches. *American Journal of Public Health*, **96** (5), 906–913.
- Pechmann, C. and Stewart, D.W. (1988) Advertising repetition: a critical review of wearin and wearout, in *Current Issues and Research in Advertising* (eds J.H. Leigh and C.R. Martin), University of Michigan Press, Ann Arbor, MI, pp. 285–330.
- Plevan, K.A. and Siroky, M.L. (1991) *Advertising Compliance Handbook*, 2nd edn., Practising Law Institute, New York, NY.
- Shimp, T.A. (2010) *Advertising, Promotion, and other Aspects of Integrated Marketing Communications*, 8th edn., South-Western Cengage Learning, Mason, OH.
- Stewart, D.W., Furse, D.H., and Kozak, R.P. (1983) A guide to commercial copytesting services. *Current Issues and Research in Advertising*, **6** (1), 1–43.
- Stouffer Foods Corp. (1994) Commission decision, 118 FTC pp. 746–820.
- Thompson Medical Company, Inc. (1984) Commission decision, 104 FTC pp. 648–844.
- Vakratsas, D. and Ambler, T. (1999) How advertising works: what do we really know? *Journal of Marketing*, **63**, 26–43.
- Winer, B.J., Brown, D.R., and Michels, K.M. (1991) *Statistical Principles in Experimental Design*, 3rd edn., McGraw-Hill, New York, NY.



Please note that the abstract and keywords will not be included in the printed book, but are required for the online presentation of this book which will be published on Wiley's own online publishing platform.

If the abstract and keywords are not present below, please take this opportunity to add them now.

The abstract should be a short paragraph upto 200 words in length and keywords between 5 to 10 words.

Author names(s) and affiliation(s) will appear, as listed below, in a List of Contributors in the front of the print edition and will also appear in the online version of the book. Please check that the names(s) and affiliation(s) given below are correct and that you (and any co-authors) are happy for this information to appear both in the printed book and in the online version.

Abstract: We define advertising copy testing, describe the major copy test methods using case examples, discuss each method's strengths and limitations, and identify areas of debate. The case examples were selected based on the authors' direct experiences and the methodological rigor of the methods. We define advertising copy testing as a quantitative and qualitative marketing research method that is used to pretest advertisements where numerical data on effects is collected and statistically analyzed. We discuss three copy test methods that apply the principles of sound copy testing that were developed by US advertising agencies, which require that the methods (i) be relevant to the objectives, (ii) have agreement about the use of results, (iii) use multiple measurements, (iv) be based on a model of human response, (v) consider multiple exposures, (vi) test comparably finished executions, (vii) control the exposure context, (viii) define the relevant sample, and (ix) demonstrate reliability and validity. To this list, we add (x) take baseline (i.e., pre-exposure) measurements and/or use control groups, because this reflects sound and generally accepted research standards. Our three case examples of sound copy test methods come from the US National Youth Anti-Drug Media Campaign, the US Federal Trade Commission, and the ARSgroup, a major copy test firm.

Keywords: advertising; copy test; pretest; evaluation; consumer research; marketing research; research validity

Author(s) and Affiliation(s):

Cornelia Pechmann

University of California – Irvine, The Paul Merage School of Business, Irvine, CA, USA

J. Craig Andrews

Marquette University, Department of Marketing, Milwaukee, WI, USA

FIRST PAGE PROOF





Author Queries

- Q1. Please spell out the first name of the author J.Craig Andrews.
- Q2. Could we change 'Method' to 'Method 1', 'Method 2', 'Method 3' at the three instances below and delete 'Copy Testing for the National Youth Anti-Drop Media Campaign
- Q3. In "Specific biases of concern are that people may purposely fully report the same beliefs at both points in time or exhibit reactance against the testing and claim adverse effects." may we change "reactance against" to "resistance to"?

FIRST PAGE PROOFS

