

EVALUATING MAGAZINE OTS IN CROSS-MEDIA ADVERTISING CAMPAIGNS

**Robert Cardarelli and William Havlena, Dynamic Logic, a Millward Brown Company
Carrie Campbell, National Geographic**

1. Introduction

Evaluation of cross-media advertising campaigns requires that the impact of each medium or media combination on awareness and persuasion be estimated in a manner that is fair to each of the media. While exposure to online advertising may be measured directly through observational methods via online surveys (for example, using Internet cookie data to track advertising exposure), opportunity-to-see (OTS), “the industry’s agreed upon measurement of audience” (Ephron, 2004) for print or other offline advertising, is often estimated based on self-reported data through the use of diaries, face-to-face interviews, online/mail surveys, etc. For the case of magazines, researchers who are evaluating the impact of integrated advertising campaigns currently use a variety of techniques, such as recent reading, frequency of reading, or specific issue readership, to estimate OTS. The present research paper evaluates these measurements not as a method for determining average issue readership (AIR) of a print vehicle, but rather explores how these methods translate to the true contribution of magazines in a mixed media schedule.

Although alternative readership measurements for estimating audiences have been extensively studied, the application of these methods in measuring cross-media advertising campaigns is less understood. An understanding of the strengths, weaknesses, and potential effect on results of alternative approaches used to measure the impact of print advertising in cross-media campaigns is necessary to make an informed decision about the best approach to use for a particular campaign. Knowledge of the influence of the measurement methodology can help the researcher choose the best alternative.

This study assesses the convergent validity of alternative measures of print OTS and how each measure relates to scores on branding metrics. The alternative measures we consider are:

- Recent reading,
- Frequency of reading, and
- Recognition of specific magazine issue covers.

We calculate the impact of alternative OTS definitions on brand metrics such as brand awareness and brand favorability. The extent to which these measures lead to similar conclusions is examined to understand the convergent validity of the measures and the sensitivity of results to changes in OTS definition.

2. Past Readership Measurement Overview

Recent reading is a well-accepted measurement that is utilized in close to 90% of the existing national readership surveys (NRS) throughout the world (Meier, 1999). Audience estimates that are based on recent reading depend on a question that “seeks to establish when a reader last read or looked through any issue of the magazine concerned” (Brown, 1990). Although an established method for estimating average issue readers, various concerns have been raised over the past twenty or more years regarding the estimates this method produces. Telescoping, the erroneous recall of the last time a title was read as being more recent than it actually was, may lead to overestimating (Val Appel 1981, p. 223). Alternatively, underestimating due to misinterpretation of how reading is understood, memory failure, or confusion over dates may also be a problem, particularly when out-of-home readership is concerned. As with most readership measurements, another factor that associated with problems related to the recent reading method is the number of titles measured in the questionnaire itself.

Frequency of reading is also a popular method used in surveys to derive probabilities of exposure. There is disparity in how the frequency (or regularity) of reading is determined, such as the types of verbal and numerical scales used. Independent of the type of scale, the frequency of reading question objective is “to divide readers into groups, with real differences between the people assigned to the different groups and, within groups, as much homogeneity as possible” (Brown, 1990). Issues that surround this method are overstated frequency, respondent confusion resulting from the use of single scales for magazines with different publication periods (Speetzen, 1988), and variability in respondents’ views on how the frequency scale translates to their perception of the publication or their actual reading behavior.

In 2003, Fasse suggested that, “specific issue readership would be a first step towards measuring the true contribution of print in a multi-media schedule” (p. 531). Showing respondents specific issue covers could help aid memory more than just a logo or masthead. As Costa Tchaoussoglou (1985 p. 410) concluded, “cover recognition seems to be a reasonable instrument to establish magazine readership, although 14% claimed to have read a future issue”. The age of the issues prompted to the respondents could present an issue on how to consider readers of a specific issue. If the issue is too young, the issue would not have an accumulated its audience, while if it is too old, some people may have forgotten long-ago reading events by the time they are surveyed.

It has been demonstrated that it is possible to develop magazine audience estimates using the Internet that generally correlate with estimates produced using a standard paper questionnaire (Elder, Incalcatera, Appel, 1999). The web administration of a media questionnaire offers many benefits, such as controlled rotation, skip patterns, and display of magazines covers in a live testing condition. However, some consequences are seen when conducting magazine measurements over the web, such as overall reduction in response rates and general increase in audience levels. MRI in 2003 has concluded that “audience estimates obtained from an economically feasible internet base data collection procedure do not mirror those based on face-to-face interviews” and the “differences in audience estimates are not constant” across titles. The purpose of the current research is not to understand how the Internet can be used to estimate overall readership levels, but rather demonstrate how the different measures translate to differences in measuring advertising effectiveness.

3. Methodology

The current research examines the incremental branding effect of advertising placed in multiple editions of National Geographic magazine across a group of six advertisers, fielded as six separate surveys by measuring the following traditional brand metrics:

- **Unaided Brand Awareness:** focuses on what brands are top-of-mind for consumers in a particular category.
- **Aided Brand Awareness:** focuses on consumers’ awareness of a brand that is presented to them by name or logo.
- **Ad Recall:** focuses on consumers’ ability to recall seeing or hearing an ad for a specific brand.
- **Message Association:** focuses on the ability of consumers to accurately link a message or statement to the brand among a competitive set.
- **Brand Favorability:** persuasion metric that focuses on consumers’ overall opinion of a brand, using a scale ranging from “very favorable” to “very unfavorable”.
- **Purchase Intent:** persuasion metric that focuses on consumers’ stating willingness to consider buying a product/brand when in-market. A variation to this is behavior intent, a gauge to measure intended behavior (e.g., likelihood to ask a doctor about a brand).

The advertisements were for companies or products in four categories: automotive, CPG, pharmaceuticals, and technology. For each of the six surveys, competitive brands within the corresponding product category (i.e. technology product) were also included for blinding purposes.

Respondents were invited to take the survey online via the National Geographic website using a web intercept technology called Dynamic Link. When a site visitor clicks on a pre-existing link within the National Geographic web page, he/she may be served a survey invitation at a predetermined sampling rate. If the respondent accepts the survey invite, the survey would then appear integrated within the webpage. Upon finishing the survey, respondents are taken to their desired destination within the National Geographic website. The sampling and completion rates were consistent across each of the six surveys fielded.

To boost the sample, email recruitment was also used. An email list of registered visitors to the National Geographic website was used to invite potential respondents into the survey, where respondents were chosen at random and sent an email to participate in the survey.

Fieldwork occurred in a “live” setting, meaning that respondents were recruited into the sample when the issue with the measured advertising was available on newsstands or to subscribers.

For each survey, readership was determined for National Geographic along with three other placebo publications within the same genre. Three versions of each survey were developed, each containing one form of the magazine readership question. After accepting the survey invitation, the respondent was randomly served a survey containing one of the following readership questions to determine OTS:

i. Recent Reading

Question: “*When was the last time you read or looked into each of the following magazines?*” The responses measured were:

- Today/Yesterday
- 2 to 3 days ago
- 4 to 7 days ago
- 8 to 14 days ago
- 15 to 30 days ago
- 1 to 3 months ago
- Longer ago
- N/A - I have not read or looked into this magazine

A respondent was considered to have opportunity-to-see if he or she read National Geographic within the period in which brand advertising appeared. For each respondent, the recent reading data was mapped back to the survey date to determine if the respondent qualifies as having OTS. For example, if a respondent was surveyed one week after the online sale where the advertising appeared, reading at least within the last seven days was required to assign OTS.

ii. Frequency of Reading

Question: *Out of the last 4 issues of each of the magazines listed below, how many have you read or looked into? (Please check one for each magazine. This does not include online versions.)*

Opportunity-to-see advertising was determined by mapping the number of reported magazines read or looked into by the date the respondent participated in the survey. For example, if an advertiser published their content in the August and July issues of National Geographic and the respondent was served the questionnaire in August, at least three out of four issues were needed to be assigned to the OTS condition. If the respondent was surveyed in August and the advertising only appeared in the August edition, then four out of the last four issues would be needed.

iii. Issue level exposure

Which of the following specific issues of magazines have you read or looked through? (Check all that apply. Skip this question if you have not read or looked through any.)



In this case, OTS was derived by respondents' recall of the specific issue(s) that they read or looked through. We measured advertising that appeared in the July and August editions. If a respondent selected an old issue that contained the advertising (before July) and did not select July or August, the respondent was then disqualified from having opportunity-to-see to control for the potential effects of advertising decay.

The probability of being served one of the above opportunity-to-see methods was set to be the same across all respondents. For each of the six advertisers measured, similar completion rates were seen across each of the OTS methods. Respondents for each study were recruited during the same time period.

For all readership questions utilized, the order of the publications (not specific issues in the issue level exposure method) was rotated to prevent any order bias. Since the number of magazines was limited to four, screening or filter questions were not required for the OTS measures utilized.

4. OTS Estimates by Measurement Type

The following table illustrates the OTS classification estimates (percentage) by measurement type:

	Classified by Recent Reading Method	Classified by Frequency of Reading	Classified by Specific Issue Exposure (Cover Images)
Advertiser #1 N=1100	83.6% n=329	57.8% n=389	65.1% n=382
Advertiser #2 N=1200	67.5% n=403	45.8% n=384	41.4% n=413
Advertiser #3 N=1140	79.8% n=391	58.5% n=369	63.9% n=380
Advertiser #4 N=1154	82.8% n=378	62.8% n=401	61.3% n=375
Advertiser #5 N=1118	79.8% n=366	50.3% n=366	46.9% n=386
Advertiser #6 N=1113	73.2% n=364	44.8% n=362	46.8% n=387

Since the survey sampling was executed on the National Geographic magazine website, the calculated levels for OTS appear to be slightly higher than they might be if they were compared to a more broad sample base (if we sampled respondents from more sites or used other traditional fieldwork methods). However, calculating a true OTS audience estimate is not the objective of the current research; instead, we are examining the degree of variation between each of the OTS assignments. These comparisons are valid, since the only difference within a survey for any given advertiser was the specific readership question administered and assignment of each question to respondents was randomized.

Using ‘recent reading’ as a measure to establish opportunity-to-see resulted in higher estimates than using frequency of reading or specific issue level exposure methods. The specific issue level and frequency of reading methods yielded similar percentages, with some exceptions, for example, within advertiser #1, respondents were more likely to report that they read the specific issue where advertising appeared than selecting the appropriate frequency threshold. Mapping recent reading data at the respondent level by reconciling the survey date with last time read resulted in a greater likelihood that a respondent would be considered to have OTS. This also may be a function of respondents being more likely to recall reading a particular publication in a shorter time period than they actually have, thus driving up the audience level, which is consistent with past research on telescoping.

Furthermore, the completion rates (number of respondents for each cell) are comparable. The total sample across the three groups for each study does not indicate any obvious pattern where one group is directionally higher than another.

5. Study Results

Now we will turn the results for one specific campaign, for demonstration purposes, as well as report the aggregated metric scores for all six advertisers.

Example 1.

The following represents data carried out for a corporate image ad campaign. Given that the focus of this research is to look at the differences between in advertising effectiveness scores across the OTS groups assigned, the metric scores are indexed to blind the actual data for the advertiser.

	Recent Reading	Frequency of Reading	Issue Level Exposure
	A	B	C
Unaided Brand Awareness	098	090	112
Aided Brand Awareness	104	098	098
Ad Recall	103	106	091
Message Association	090	094	116
Brand Favorability	109	096	095
Purchase Intent	108 ^B	088	103
<i>Sample Size</i>	<i>272</i>	<i>176</i>	<i>171</i>

A/B/C statistically significant at 95% confidence level.

Differences were tested with ANOVA at the 95% confidence level (assuming equal variances). Metrics producing significant results were tested with pair wise t-tests to identify specific differences.

Overall, there were few statistically significant differences across the three measurement types. The superscript ‘B’ indicates that, in this situation, respondents classified as having OTS by recent reading showed significantly higher purchase intent than those classified using the frequency of reading method.

Aggregate data for 6 campaigns:

The following table represents the aggregate averages of the six campaigns

	Recent Reading	Frequency of Reading	Issue Level Exposure
Unaided Brand Awareness	28.4%	26.7%	29.0%
Aided Brand Awareness	87.2%	87.4%	87.2%
Ad Recall	50.0%	50.5%	48.1%
Message Association	12.5%	11.7%	12.3%
Brand Favorability	57.6%	57.2%	56.9%
Purchase Intent	47.7%	43.7%	45.6%
<i>Sample Size</i>	<i>1729</i>	<i>1216</i>	<i>1255</i>

In the aggregate, the results across the six advertisers indicate that there are relatively small differences in how each of the measurement techniques relate to differences in brand metric scores. For each of the six studies there were six metrics, resulting in 108 possible comparisons. Out of the 108 possible comparisons, there were only 6 significant differences observed. Respondents assigned to OTS by recent reading scored significantly greater than the other methods in 5 out of those 6 times (once for unaided brand awareness, ad recall, and message association and twice for purchase intent). Regarding the other significant difference, specific issue level reading OTS was greater than frequency of reading for one advertiser. This case could be attributed to random sampling error, given the number of possible cells tested.

6. Conclusions

The current research has demonstrated that the specific method for assigning OTS has marginal impact on the overall results when measuring advertising effectiveness. Using recent reading to classify respondents tends to produce higher overall estimates (number of respondents that qualify). However, this translated to statistically significant differences in ad effectiveness metrics for only 5 out of the 108 tested combinations.

When measuring print OTS in a cross media advertising effectiveness study, the case has been made that using specific issue cover images is the best method to assign OTS. However, using a frequency of reading scale to map respondents based on date produces very similar results as the specific issue approach. When measuring integrated campaigns that have a wide distribution of print vehicles, having too many cover images on the computer screen may affect survey participation rates. For example, too many images may slow down the speed at which the survey page loads. Respondents may be overwhelmed by having to evaluate too many issues per page, which may increase respondent fatigue by asking him or her to recall every specific issue read. In some cases, using a frequency scale may be more appropriate to determine OTS, particularly when working with advertisers or agencies that change their plans in-flight. In this case, moving the OTS calculation to the back end of the data processing design would allow the researcher to be more flexible when updating the OTS assignment.

References

- Appel, Valentine (1981), "Telescoping: The Skeleton In The Recent Reading Closet," Proceedings of the 1981 Worldwide Readership Research Symposium, Session 2.17, England: Ipsos-RSL, 223-226.
- Brown, Michael (1990), "Dear Reader, Some readership measurement questions...and some answers", Research Services Ltd and British Market Research Bureau, p. 37,126
- Edler, Andrew, Tony Incalcaterra, and Val Appel (1999), "Measuring Print Audiences Via The Internet," Proceedings of the 1999 Worldwide Readership Research Symposium, Session 9.4, England: Ipsos-RSL, 509-516.
- Ephron, Erwin (2004), "Sauce for the Outdoor Goose," Ephron on Media Newsletter, January 9
- Faasse, John, Leendert van Meerem, (2003), "What the World Needs Now," Proceedings of the 2003 Worldwide Readership Research Symposium, Session 7.4, England: Ipsos-RSL, 531-538.
- Frankel, Martin, Julian Baim, Michal Galin, Michelle Leonard, (2003) Proceedings of the 2003 Worldwide Readership Research Symposium, Session 3.3, England: Ipsos-RSL, 131-141.
- Meier, Erhard (2003), "What's Happening Where?" Proceedings of the 2003 Worldwide Readership Research Symposium, Opening Session Harrow, England: Ipsos-RSL, 1-8.
- Speetzen, Rolf (1988), "Current Trends In The Work Of The AG.MA," Proceedings of the 1988 Worldwide Readership Research Symposium, England: Ipsos-RSL, 463.
- Tchaoussoglou, Costa (1985), "Ad Recognition And Cinderella's Lost Slipper," Proceedings of the 1985 Worldwide Readership Research Symposium, England: Ipsos-RSL, 410.

