

PRINT ADVERTISEMENT MEASUREMENT: GETTING INTO THE NITTY GRITTY

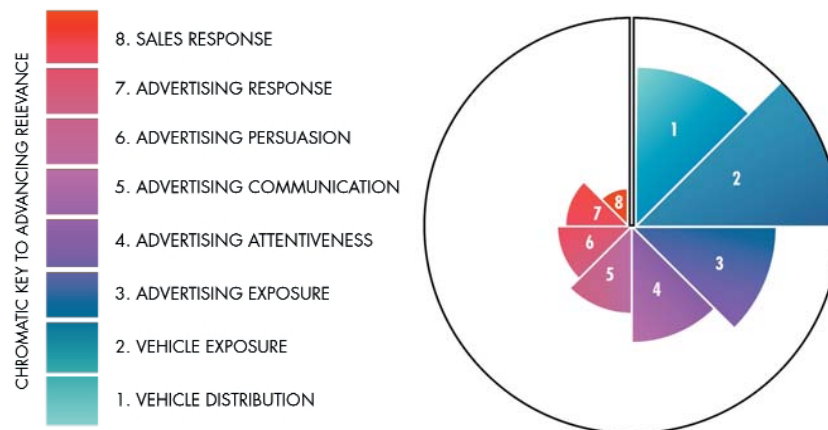
Michal Galin, Julian Baim, Martin Frankel, Konstantin Augemberg, and Valerie Veith
Mediamark Research & Intelligence

Introduction

The history of media measurement is dominated by offerings of audience level data that generally leave users (buyers and sellers) at the level of “opportunity to see” an advertisement. This information has been integral in dealings and negotiations between parties involved in the buying and selling of print advertisement space. The ability to accurately and consistently size audience and describe the profile of this audience is therefore imperative in the campaign planning process. Most importantly, it allows parties to select appropriate vehicles for their product messages.

But vehicle audience/exposure is only one dimension in the media planning and buying dynamic. And, this dimension does not allow the discussion to effectively answer questions about the advertisement such as: How many people did my ad actually reach? What was the return on the investment? Did my ad work?

Ultimately, advertisers need to have information pertaining to both their particular ad and their full campaign. They need to have information so that they can evaluate their return on investment (or even objective). The figure shown below is an ARF model published in 2003. As Faasse asserts in this 2007 paper, “this is a good starting point for a systematic analysis of the measurement problems we are facing today.” (Faasse, 2007) The important take home message of this model is that media must ascend the hierarchy into more granular data in order to maintain/increase relevance.



Such ad related measurement and data are not new, as further demonstrated by the references reviewed in preparation for this paper. But, most often, until recent time these research endeavors have been isolated and ad hoc. With intensifying calls for more granular accountability media metrics and more transparency, ad related measurement has moved into the realms of syndicated and comprehensive approaches.

These syndicated studies of ad effectiveness allow users to explore a multitude of variables that may contribute to the overall success of an ad. There are many elements of an ad’s DNA that may interact to determine its ultimate success: elements inherent to the ad itself (e.g. size, color, positioning in the magazine, etc.), the fit/interest between the readers and the product, and the



actual creative of the ad. And, it is important for researchers to evaluate all of these elements to understand what drives the performance of an ad. These analyses should also inform discussions of return of investment/objective in that they impact these ultimate evaluations. The focus of this paper is on analyzing some of these ad variables: size/color, positioning and elements incorporated into the actual ad.

Background – Starch Measurement

In his 1923 book entitled Principles of Advertising, Daniel Starch laid out the following:

The functions of an advertisement are fivefold: To attract attention (the advertisement must be seen); to arouse interest (the advertisement must be read); to create conviction (the advertisement must be believed); to produce a response (the advertisement must be acted upon); and to impress the memory (the advertisement in most instances must be remembered). (Page 7)

MRI Starch has been collecting data on print advertisement effectiveness since 1923. Up until the about 2004, these data were collected primarily through in person and mail surveys. With the advent of the Internet and its proliferation of use as a research tool, Starch began to utilize this mode as a means to more efficiently and effectively extend and supplement the measurement of advertisements.

Among other information, the Starch methodology collects the following key readership metrics on each ad: noting, associated, read some/most. The noting score is the quintessential score indicating what percent of readers remembered seeing the actual ad. This is the information that moves the “opportunity to see” needle to the “actually saw” position. This information offers a response to the accountability demands of advertisers and agencies. Noting scores will be used in many of the comparisons shown below to explore differences in various elements of ad DNA. (Noting scores are represented as a percent.)

Of those readers who noted the ad, Starch delves a bit deeper to ascertain whether they knew what brand was being represented in the ad (the associated score) and, where applicable, whether the reader engaged further with the ad by reading the text included in it (the read some/most scores).

Beyond the data collected in Starch surveys, each ad examined by these surveys is categorized along other variables: size/color, positioning within the book, and more recently, actual aspects of the ad (if it incorporates a website address, a model, a high impact element, a coupon, a recipe or other similar elements). It is the melding of these ad variables and survey data that allows researchers to explore some key determinants of ad success.

Advertisement Databank Description

Much of the discussion in this paper focuses around the analysis of 12,570 advertisements measured across 270 issues of 112 magazine titles. The time period of this measurement occurred between October 2008 and April 2009. In many cases, the analysis is broken out by particular magazine genre, type of ad size and/or product category.

Print Advertisement Insights

Position Analysis

In Fielding and Bahary’s 2003 Worldwide Readership paper, much attention was paid to the question of ad positioning (Fielding and Bahary, 2003). Fielding and Bahary conducted primary research among 24,000 households across a four week period. The



research involved showing respondents ads and asking them follow-up questions including whether they recalled seeing the ad. This study found that “positioning does have a significant impact on the recall of advertising.”

Analysis of Starch data yields similar findings:

- Ads on covers tend to perform better than ads within the magazine. There is additional discrimination among ads in different cover positions (i.e. second, third, fourth) within the magazine,
- There is some indication that ads in the front of the magazine perform better than ads in the back of the magazine, in particular within magazine genre.

Examining all ads measured from October 2008 to April 2009, we find that second covers, in particular, tend to perform better than any other position in the magazine. This is good news for advertisers who generally pay higher prices for placing their ads in this premium position.¹ The position analysis across all titles included in this analysis can be found in Table 1 below.

Table 1: Position Analysis Across All Measured Titles (1-page 4 Color Ads: 7699 ads)

Position	Noted Score (%)	Associated Score (%)	Read Some Score (%)	Read Most Score (%)
Second Cover	66	57	57	23
First Quarter of Book	50	44	40	18
Second Quarter of Book	47	41	38	17
Third Quarter of Book	46	40	37	16
Fourth Quarter of Book	45	38	36	16
Third Cover	56	46	47	19
Fourth Cover	59	54	49	21

¹ It should be noted that second covers are calibrated within the Starch database. Starch’s long and rich history of data collection enables comparison of ad noting scores (and other measures) between surveys conducted in-person and online, respectively. The analysis showed that in-person and online methodologies produce consistently different ad noting and other effectiveness scores for certain types of ads. In particular, Starch’s analysis suggests online surveys cannot adequately reproduce the same context, texture, and form of certain ad types as when interviews are conducted in-person. As a result of having a large, historical database, MRI Starch has been able to establish calibration factors to adjust ad effectiveness scores for these ad types so that they better reflect scores obtained from in-person surveys. Along with second covers, the ad types impacted by this calibration are: inserts, gatefolds, and multi-page ads. Ad noting and related measures are therefore calibrated before the release of the data. Our analysis also indicates that other ad types are not impacted by different survey modes.



Third and fourth covers also tend to perform better than other ads in a magazine, but do so at a lower rate than second covers. The relative strength of ads on covers is found across all magazine genres (again, with second covers performing the best across genres).

But, there are a limited number of cover ad opportunities within an issue and advertisers may not be willing or able to pay for these premium positions. So, how do ads within the issue tend to behave? Are there insights that we may draw from analyzing front of book versus back of book data? And, are there differences when one compares different magazine genres?

The answer to all of these questions is generally yes.

When one examines the data across different magazine genres, it is clear that there are differences in ad performance between ads shown at the beginning of the magazine versus the end of the magazine. Starch breaks magazines into 17 genres or categories. The data within book position are shown in Table 2.

Table 2: Position Analysis (1st Quarter vs. 4th Quarter) By Magazine Genre (1-page 4 Color Ads: 7699 Ads)

Magazine Genre	Average Noting Score (%)– 1 st Quarter of the Magazine	Average Noting Score (%) – 4 th Quarter of the Magazine
Automotive	53	44
Bridal	51	49
Business/Finance	49	39
Entertainment	49	42
Epicurean	51	46
Family/Parenting	51	48
General Interest	53	50
Health/Fitness	51	48
Home/Home Service	49	47
Men's	49	42
Music/Music Trades	52	46
Science/Technology	50	39



Sports	48	44
Travel	51	40
Weekly News	44	42
Women's Fashion/Beauty	53	47
Women's Service	48	46

(Differences are significant at the $p < .01$ level)

Interestingly, ads placed in the back quarter of the book in the automotive category, on average, yield noting scores that are 9% lower than ads placed in the beginning quarter of the book. For the business/finance category, this differential is 10%. And, for the travel category as well as the science/technology category, the difference between placing an ad in the front quarter of the book versus the back quarter of the book translates into a difference, on average, of about 11% in noting score.

While there are categories that see some notable differences in noting scores in the front versus back of the book, there are other categories where the differences are not great at all. For example, there is only a 2% differential, on average, between noting scores in the women's service category, in the home/home service category and in the weekly news category. Of course, when translating these percentages into actual audiences, as MRI Starch does with Ad Measure, even these small differentials can indicate large losses of audience.

Another aspect of ad positioning that we examined is the placement of ads either on the right hand side of the magazine or the left hand side. Are there any differences in the data that would compel advertisers to select one side over the other? The short answer to this question is no. Looking across every one of the standard Starch readership metrics, we see no difference in performance of ads placed on either side of the magazine (Table 3).

Table 3: Right Hand/Left Hand Position Analysis (1-page 4 Color Ads: 7699 Ads)

Position	Noted Score (%)	Associated Score (%)	Read Some Score (%)	Read Most Score (%)
Right Page	48	42	39	18
Left Page	47	41	38	17

It is important to note that this information uses data from magazine measurement in the United States exclusively. Interestingly, Foley found in his 1999 Worldwide Readership Symposium paper that for newspaper measurement in Asia there was some difference in the performance of ads placed on the right versus left hand side of the page. In Foley's work, the impact of reading direction (left to right versus right to left) explained the differences in findings for Hong Kong versus Singapore.

Position in the magazine does seem to have some influence on the success of an ad, in particular for some genres. This finding does not extend to placement of the ad either on the right or left side of the magazine.

Size of Ad Analysis



Daniel Starch’s book also discusses the concept of ad size as it pertains to impact and return on investment. Much of his discussion focuses on assessing the benefits of larger ads with regard to performance/recall metrics by the accompanying incremental cost of producing such ads. Ultimately, he concludes that:

...large space makes a more intense impression by its sheer magnitude...has less competition with other advertisements for the reader’s attention...is able to secure the reader’s attention more exclusively...permits more adequate presentation of the proposition...and tends to create an impression of greater importance and reliability of the firm which is advertised. (Page 578)

While in his book Starch was really referring to full-page ads versus half-page or even smaller ads, our more modern world includes a panoply of ad sizes that were probably not dreamed of in 1923. While 1-page 4 color ads are still probably the most common ads, we now have examples of multi-page ads of upward of 11-pages or more -- there are gatefold ads, insert ads and booklets. The types of ad sizes have grown and diversified in an attempt to create new impressions on readers.

But does size matter? Do larger size ads perform differently from their smaller cousins?

The simple answer is that there are some examples indicating that size does matter. We first analyzed the data for more commonly sized ads: 1-page 4 color ads versus 1-page 4 color spreads. Overall, spreads (1312 ads analyzed) perform slightly better in noting versus their smaller, one-page (7699 ads analyzed) counterparts (53% versus 47%, respectively). Delving even deeper into a comparison of spreads versus one-page ads on a category basis, one finds some interesting stories (Table 4). Information such as this can be utilized to inform decisions made by advertisers and agencies when creating and placing ads. Additionally, these data may be analyzed by title and magazine genre for further insights. (That analysis is outside the purview of this paper.)

While most product categories really show no difference in ad noting performance in one-page ads versus spreads, there are a few notable exceptions. For example, looking at the Cosmetics & Beauty Aids category one can see that spreads tend to produce, on average, 9% higher noting scores. And, in the Automotive & Automobile Accessories category there is a difference, on average, of 7% between one-page ads and spreads. And while there are not an abundance of examples of liquor spread ads, there is some indication in the data that spreads in this category perform better than one-page ads. The Jewelry & Watch category is another interesting example where spreads perform noticeably better than one-page ads (48% noting score for one-page ads versus 59% noting score for spreads). Again, there are not many examples of spreads in this category but the data suggest that the larger units achieve a sizeable difference in noting scores.

Interestingly, with the series of examples presented in this table, there is not one example that shows spread units performing less well than one-page units. Either these two types of ads produce very similar levels of recall by magazine readers or spreads generate higher metrics. It is important to note that this analysis simply focuses on the size of the ad. There are probably a multiple of variables interacting to produce the results that we ultimately see. Such variables include the placement of an ad within a particular vehicle and the actual creative aspects of the ad.

Table 4: Noting Score Comparison by Size Within Product Category

Major Product Category	Number of 1-page 4 Color Ads	Number of 1-page 4 Color Spreads	Average Noting Score (One-Page)	Average Noting Score (Spread)
Financial	275	31	44	45
Jewelry & Watches	218	18	48	59



Computers, Software, Internet	161	26	43	51
Cosmetics & Beauty Aids	446	112	51	60
Hair Products & Accessories	121	46	50	57
Prescription Medications	333	52	35	36
Automotive & Automobile Accessories	288	106	49	56
Retail	423	140	48	53
Prepared Foods	295	10	55	57
Liquor	90	13	50	59
Resorts & Travel Accommodations	270	66	47	50

(Differences are statistical significant at the $p < .05$ level)

There are a multitude of product categories to choose from, but this discussion focuses on examples that tend to have larger numbers of ads available for analysis and comparison. The product categories were also selected for inclusion in this paper based on comparable data in both size designations.

Turning our attention to the findings for multi-page ads, one can see that larger size ads perform marginally better as they grow in size. However, this difference is more pronounced as compared against the standard of 1-page 4 color ads; one can see a very large difference in performance (47% noting score for 1-page ads versus 72% noting score for 8-page ads). Larger sized units seem to capture the attentions of readers more effectively than smaller sized units. These data support Starch’s claims about the value of larger-sized ad units.

Table 5: Noting Score Comparisons for Multi-Page Ad Units

Ad Size Designation	Number of Ads Available for Analysis	Average Noting Score (%)
1-page 4 color	7699	47
2 page 4 color	84	54



3 page 4 color	70	63
4 page 4 color	97	66
5 page 4 color	10	64
6 page 4 color	27	71
8 page 4 color	12	72

In comparing other types of special ads, we see a similar pattern. The larger the ad, in general, the higher the noting scores. For example with the case of gatefold ads; the bigger the size of the gatefold, the higher the noting score. Table 6 displays information for two types of gatefold ads -- gatefold inserts and standard gatefolds.

Table 6: Noting Scores for Gatefold Ads

Type of Gatefold	Number of Ads Available For Analysis	Average Noting Score (%)
1-page 4 Color Gatefold Insert	6	41
3 Page 4 Color Gatefold Insert	6	63
4 Page 4 Color Gatefold Insert	15	67
2 Page 4 Color Gatefold	5	51
3 Page 4 Color Gatefold	5	56
4 Page 4 Color Gatefold	7	65

At present, we do not yet have a multitude of examples of gatefold ads. The database, however, is continuously building inventory of these ads types. We can still see a clear directional story that the bigger the gatefold size, the higher the noting score. While these ads are probably higher in cost to produce, they do result in increased levels of attention grabbing.

Since we saw earlier in our analysis that second covers tend to perform better than other positions, we extended our analysis of ad sizes by examining their relative performance on second covers.

Table 7: Comparing Differently Sized Cover Ads



Second Cover Ad Description	Number of Ads Available for Analysis	Average Noting Score (%)
Second Cover 1-page 4 Color	55	66
Second Cover 4 Color Spread	167	67
Second Cover 3 Page 4 Color Gatefold	5	75
Second Cover 4 Page 4 Color Gatefold	19	76

While the number of examples of the larger type second cover ads is obviously limited, there is some indication that those larger sized ads outperform the more standard second cover sized ads (one-page and spreads).

Additionally, we analyzed the other side of the size continuum. Do ads smaller than 1-page 4 color perform worse, in general, to their larger cousins?

The answer seems to be yes. In looking at both half page ads and one-third page ads, one can see that the noting scores tend to be lower, on average, than 1-page 4 color ads.² While 1-page 4 color ads have an average noting score of 47% across all measured ads (n=7699), the comparable figure for half page ads is 38% (n=7) and for one-third page ads 37% (n=12).

These data beg the question: Do the ad performance data comparisons shared in this section of the paper warrant the added cost of creating and running larger ads? The response to that question, while critical to the industry, is outside the purview of this paper.

A Bit on the Interplay Between Size and Positioning

We further continued the investigation on the performance of ads taking into consideration the combined impact of their positioning in the magazine and their size. Additionally, we infused product category into this analysis to provide additional insight. While this analysis is interesting the small numbers of ad examples that are currently available ultimately limits it.

Some of the findings from this analysis of variable interplay can be found below.

- While spreads and one-page financial ads produce very similar noting scores (45% versus 44%), second cover financial spreads produce a 61% noting score (n=7 ads). That is close to a 20% lift in noting scores.
- While there is some difference in how spreads versus one-page automotive ads perform, comparing the data to second cover spreads shows an additional lift for ads within this category. The average noting score across the 48 second cover automotive spreads was 67% -- this is 11% higher than noting scores for automotive spreads (56%). Additionally, we see that third and fourth cover automotive ads (one-page) perform at similar levels to automotive spreads (54% and 56%, respectively).
- Extending these comparisons to the Cosmetics & Beauty Aids category we see similar results. Second cover-spread ads (n=18) in this category succeed in lifting average noting scores to 74% (versus 51% average noting scores for one-page ads and 60% average noting scores across spreads). Again, we see that third and fourth cover cosmetic ads do not

² Note: The Starch methodology measures all nationally running ads of one third of a page or larger.



yield the higher level of noting scores that second covers produce. Third (n=10, average noting score 54%) and fourth (n=40, average noting score 62%) cover ads in this category tend to behave more like one-page ads and spreads.

Evaluating the Impact of Ad Characteristics on Ad Performance

We extended our analysis to variables beyond the general characteristics of the ad to actual elements incorporated into the ad itself. In order to distinguish these elements from the types of characteristics discussed above, we have defined those elements discussed above (ad positioning and ad size), as level one characteristics. The elements used to focus the remainder of this discussion will be called level two characteristics (presence of coupons, nutritional information, scent strips, recipes, etc. within the ad). These level two characteristics are not included in all ads. And, there are many ads that include multiple level two characteristics. These more granular characteristics allow us to delve deeper into qualities of a given ad that may be contributing to its overall performance. Our goal was to produce insights that may help those who create ads. Two such analyses were conducted and will be discussed in this paper: (1) food ads and (2) high impact ads.

(1) Food Ads

Our driving objective for this analysis was to evaluate if different aspects of food related ads influenced Starch standard readership metrics in the positive direction.

We isolated 972 food ads (ads related to any food categories) measured between October 2008 and April 2009 to be hand coded along several variables. The list of elements that each ad was reviewed for: coupons, recipes, nutritional information, sweepstakes, freebies/sample offers, and website address. Our coding exercise found that the most common of the characteristics in this set of ads were: ads with recipes (17.5% of all ads coded), ads with nutritional information (28.2% of all ads coded) and ads with website addresses (12.4% of all ads coded). As a point of reference, the food categories that were represented in this project were extensive. The food categories for which the most examples were found: bakery goods, beverages, cereals, dairy products, prepared dinners and entrees, soups and snacks.

The food ad characteristic that we found to contribute most positively to the performance of such ads was the presence of a recipe. Across the board, comparing all Starch metrics, food ads with recipes tended to yield higher scores than food ads that did not include recipes. For example, the average noting score for food ads with recipes was 62% versus the average for ads without recipes of 53% (this is statistically significant at the .05 level). The same positive results hold true for the other Starch metrics (associated, read some/most) in comparing food ads with and without recipes.

Furthermore, turning to a Starch metric we have not yet discussed in this paper, actions taken as a result of seeing the ad, food ads that incorporate a recipe saw significantly higher “clipped the ad” rates than ads without recipes (10% versus 3%). Again, this is statistically significant at the .05 level. Overall, we found that ads with recipes result in statistically significantly higher “any action taken” scores versus ads without recipes.

Analysis around the other prevalent food ad characteristics, nutritional information and website addresses, did not reveal any significant differences between ads with and without such characteristics. And, it should be noted, that for some of the coded characteristics, there were not many examples found so the analysis was limited.

Finally, one can also evaluate whether ad size patterns that hold true generically across ads or by magazine genre also apply to ads within a product category. We found that 1-page 4 color food ads had average noting scores lower than those for 1 page spreads (54% versus 61%). And, evaluating food ad inserts we see a similar pattern. As the insert increases in size from a 1 page insert to a 4 page insert, so does the average noting score for those ads (48% versus 78%).



(2) High Impact Ads

The main question driving this analysis stemmed from the desire to scrutinize whether ads with some experiential element performed better than ads without these special elements.

For this analysis, we reviewed all ads within three magazine genres: women's service, women's fashion and epicurean. We reviewed close to 5,500 ads for this particular analysis across 34 issues measured between October 2008 and May 2009 in those three magazine genres. In the case of this analysis, we coded the presence of the following ad characteristics: coupons, inserts, gatefolds, scent strips, business reply card, glue-ons, perforated ads and sponsored editorials. It should be noted that we did not find a tremendous number of examples of ads with such characteristics within the 34 issue set. The characteristics with the greatest numbers of examples are as follows: coupons (3% of all ads coded), inserts (2% of all ads coded), and gatefolds (2% of all ads coded).

We found that gatefold ads performed significantly better than ads without gatefolds across these three magazine genres (56% versus 48%, respectively). This is consistent with the findings presented earlier across all magazine genres.

Additionally, we found some interesting results for ads with scent strips. While we did not find too many examples of these types of ads (n=29), when we compared the Starch metrics for scent strip ads versus non-scent strip ads we found that the former yielded statistically significantly higher overall scores (57% versus 48%, respectively).

We will continue to code ads along such characteristics and others in order to build up our pool of examples from which to conduct analysis. We hope that as the dataset grows, we will be able to conduct more robust analyses and produce more stable insights.

Summary

The discussion driving this paper was intended to demonstrate the variety of data points available for analyzing print ads. The great news is that we have so much data and so many opportunities for analysis.

We found that size and positioning does matter in some cases. And, when delving into an analysis around the contents of an ad, we did see some evidence that different characteristics may have a positive impact on the overall performance of an ad.

Our future work will delve deeper into explorations of what drives ad performance.

References

Bailey, Jane, Caryn Klein and Craig Gugel. "Secrets to Success: Real-World Relationships Between Print Effectiveness, Readers & Advertising." Paper presented at the Worldwide Readership Symposium, Prague, Czech Republic: 2005.



Faasse, John. “The Virtual Currency.” Paper presented at the Worldwide Readership Symposium, Vienna, Austria: 2007.

Fielding, Richard and Judy Bahary. “Now You See Me, Now You Don’t! Does Ad Positioning Matter?” Paper presented at the Worldwide Readership Symposium, Cambridge, Massachusetts: 2003.

Foley, Tim and Steve Garton. “When Late Left Beats Early Right.” Paper presented at the Worldwide Readership Symposium, Florence, Italy: 1999.

Lurin, Mitch. “Editorial Targeting (Or, Improved Advertising Positioning In Magazines).” Paper presented at the Worldwide Readership Symposium, Barcelona, Spain: 1988.

Starch, Daniel. Principles of Advertising. A.W. Shaw Company, 1923.