



University of
South Australia

Ehrenberg-Bass
Institute for Marketing Science

Documenting patterns in advertising post- testing measurement

Emily Gray

BACHELOR OF BUSINESS
(MARKETING)

SUPERVISION TEAM:

Dr. Nicole Hartnett

Prof. Rachel Kennedy

Dr. Zachary Anesbury

Ehrenberg-Bass Institute for Marketing Science
University of South Australia

MARCH 2022

Declaration

I declare that this thesis presents work carried out by myself and does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; to the best of my knowledge it does not contain any materials previously published or written by another person except where due reference is made in the text; and all substantive contributions by others to the work presented, including jointly authored publications, is clearly acknowledged.

This research is supported by an Australian Government Research Training Program (RTP) Scholarship

.....

Emily Gray

June 2022

Acknowledgements

Thank you to my supervision team, Nicole, Rachel, and Zac for your knowledge, and guidance over the two years. To Kirsty, thank you for your support, the chats, and laughs.

To my fellow HDR colleagues: Ruby, Chandler, Brooke, Anna, Shreya, Alicia B and G, Jake, and even Stan. The chats, coffees, and laughs over the two years made this journey easier.

A big thank you to my family and friends for understanding and supporting me while I complete my masters. Looking forward to finally catching up now that this chapter has come to an end.

To my partner Kate, you are the best finding in my master's journey. Thank you for your constant support and encouragement along the way. I could not have done this last year without you. Finally, to our dog, Parker, thanks for always bringing me joy between writing sessions.

Abstract

Post-testing surveys provide advertisers measures of consumer response to advertising and are used to evaluate campaign performance. Performance of campaigns is evaluated against benchmarks that are specific to a market research provider's databank and/or the advertiser's conditions. This thesis independently reviews post-test survey data to better understand how ads perform across a large collection of intermediate measures, how scores vary across conditions and explores the relationship between these measures.

POST-TEST SURVEY SCORES WITHIN AND ACROSS MEASURES

The variation in intermediate post-test survey results is an underexplored area. Previous research either uses small ad samples of one or two ads (e.g., (Carter, Donovan, & Jalleh, 2003; Heath & Nairn, 2005; Stewart, Farmer, & Stannard, 1990) or does not detail the measurement approach (i.e., question formats, wording, prompts, response sets, and metric calculations) (e.g., (Batra, Lehmann, Burke, & Pae, 1995; Putte, 2009; Stapel, 1994). In addition, not all academic publications state measures of dispersion such as range, lowest scoring ad, highest scoring ad, or standard deviation (e.g., (Bergkvist & Rossiter, 2008; Greene, 1992; Haley & Baldinger, 1991; Kennedy, Sharp, & Rungie, 2000). These factors limit the understanding of how ad scores may vary within intermediate measures.

Further, across the industry there are different measurement approaches including different question formats, types of prompts, response sets, and calculations to determine key metrics. There is little knowledge on how ad scores collected by different approaches for the same measure may change the means or distribution of scores. With little to no public knowledge on how and/or if such differences change the scores (i.e., means, median, modes) it is hard for marketers to understand if and when ad scores are comparable.

POST-TEST SURVEY SCORES ACROSS CONDITIONS

Researchers have explored the influence of different conditions such as product category (Biel & Bridgwater, 1990; Rossiter, Percy, & Donovan, 1991; Sharp, 2007), brand size, (Romaniuk, Sharp, Paech, & Driesener, 2004; Sharp, Beal, & Romaniuk, 2001, 2002a), and ad length (Bellman, Kennedy, Varan, & Nenycz-Thiel, 2020; Newstead & Romaniuk, 2010; Singh & Cole, 1993), on post-test scores. While there are some comparisons of ad scores across conditions, there is a lack of current documentation on how scores change across conditions.

THE RELATIONSHIPS BETWEEN INTERMEDIATE MEASURES

Researchers have identified some relationships between key measures (du Plessis, 1994a; Greene, 1992; Kennedy & Romaniuk, 1999; Putte, 2009; van Kuilenburg, de Jong, & van Rompay, 2011; Venkatraman et al., 2015), contradictory findings are also evident.

METHOD AND DATA

This research reports on real-world data and a Many Sets of Data (MSoD) approach paired with statistical testing to understand post-test survey results. This thesis includes five diverse data sets containing over two thousand ads aired across nine categories in five countries. The analysis of eighteen advertising measures, including three memory, eleven evaluation, one ad communication, and three claimed-behavioural measures. Due to inconsistencies in measurement approaches, not all measures were explored across all conditions.

KEY FINDINGS

RQ1) To what extent do post-test scores vary within and across measures?

Memory measure scores (i.e., ad awareness, ad recognition, and brand linkage) have the highest levels of variation than any other measure classification. From the lowest to the highest scoring ads, ad awareness scores have the lowest variation (53pp) while brand linkage scores have the highest variation (99pp). The high levels of variation in memory measure scores may make it harder for marketers to predict target points for their campaign objectives.

Evaluation scores vary from 42pp (relevant) to 87pp (entertainment). Ad communication scores vary from 58pp (message comprehension scores collected with negative worded item) to 96pp (message comprehension scores collected with positive worded item). Claimed behavioural measure scores vary 40pp for both shopper intentions and action intentions, and 56pp for purchase intentions. There are more constraints when marketers set target points for their campaign objectives when evaluative type measures have lower levels of variation.

It is apparent that providers use different measurement approaches to collect data for the same measures, that could contribute to differences in the variations observed in the overall scores. This justified analysing the data across measurement approaches and led to a secondary investigation of *how do different measurement approaches affect variation for the same measure?*

This thesis finds that scores collected with different question formats, prompts, wording, response sets, and metric calculations changes the central tendency and distribution of scores. Therefore, marketers cannot compare scores across campaigns, regions, time, or brands if the data was collected with the different approaches.

RQ2) To what extent do post-test scores vary by conditions (e.g., product category, brand size, and ad length)?

Only subscription markets have similar scores. While there were no identifiable patterns, there are signs that categories operating in subscription markets (e.g., banking, telecommunications, tax services) are similar. For marketers and providers, category-specific benchmarks are necessary. Marketers managing brands across categories (i.e., Virgin – airlines, telecommunications, and media) cannot directly compare ad scores across their portfolios to judge ad response.

Larger brands have higher ad awareness scores. In line with previous findings (Romaniuk et al., 2004; Sharp et al., 2001, 2002a), ad awareness scores collected for large brands have higher means than medium and small brands. This indicates that marketers managing multiple brands need to account for the size of the brand when interpreting ad awareness scores.

Longer ads have higher ad recognition and likeability scores. Ad recognition and likeability scores consistently show that 60-second ads have higher means than 15 and 30-second ads. This research supports prior knowledge that 15-second ads are almost as effective as 30-second ads in ad recognition, brand linkage, and likeability (Bellman et al., 2020; Newstead & Romaniuk, 2010) while longer ads have diminishing returns (Bellman et al., 2020; Singh & Cole, 1993). For marketers, these findings suggest benchmarking needs to be like-for-like for 60-second ads but not between 15- and 30-second ads.

RQ3) How do post-testing measures interrelate?

Likeability is the measure most consistently related with other intermediate measures. First, this thesis finds supportive evidence of previous research indicating that likeability has a strong positive relationship with ad recognition (van Kuilenburg et al., 2011). This may be because ads that are liked are given more mental processing (Biel, 1998). Second, there is a strong positive relationship between likability and message comprehension (positive worded items). This is a novel finding not previously reported. This thesis finds that likeability has a strong positive relationship with entertainment, providing contradictory evidence to (Greene, 1992). Additionally, likability has a relationship with claimed attention, which has not been previously documented. Overall, likeability may be a good predictor of total ad performance.

Five other relationships were identified in this research: entertainment and claimed attention; appealing and claimed attention; claimed attention and identify with the ad; entertainment and identify with the ad; and appealing and identify with the ad. The findings suggest that these scores capture the same information which can help marketers understand possible ad performance scores: If the ad does well on one measure, it should typically score well on the other. If it does not, then these measures are starting points to explore when scores do not meet required targets.

SUMMARY

These empirical findings provide new understanding of post-test survey scores with real-world data. The information gained from this research provides marketers with an understanding of when scores can be compared and when they cannot. Along with understanding the relationships between measures, they may help explain ad performance. This helps to ensure that marketers and brand owners make fair assessments based on informed answers on whether ad campaigns have reached realistic targets.