Assessing Progress In Advertising-Related Scale Development & Usage

Gordon C. Bruner II, (E-mail: gcbr@cba.siu.edu), Southern Illinois University
Maryon F. King, (E-mail: mking@cba.siu.edu), Southern Illinois University

Abstract

An awareness in the advertising industry of the inadequacy of current practices in copy testing led to the development of PACT (Positioning Advertising Copy Testing), a statement of fundamental copy testing principles developed by twenty-one of the world's leading advertising agencies. A primary concern of this coalition was that too much reliance had been placed on single-item measures to assess the performance of ads, with a concurrent lack of validation of the measures being used. Subsequently, few studies have followed up on the PACT recommendations, resulting in little in the way of objective data to assess the progress made in the development and use of multi-item scales. This study offers an initial examination of the progress made in response to the PACT recommendations by developing an inventory and evaluation of advertising-related scales. Insights from this analysis lead to specific observations regarding the state of the art in advertising research as well as recommendations for future scale research.

1. Scale Development and Usage

In the early 1980's there was an awareness in the advertising industry that the current state of copy testing was inadequate (Yuspeh 1982). In an effort to address the situation, twenty-one of the world's leading advertising agencies initiated extensive discussions regarding copy testing issues and practices. The culmination of these discussions was the release of PACT (Positioning Advertising Copy Testing), a statement of fundamental copy testing principles. Among other conclusions drawn by this coalition, it was determined that too much reliance had been placed on single-item measures to assess the performance of ads, and that there had been a lack of validation of the measures being used.

In a recent critical review of the state of advertising research it was observed that there have been too few studies following-up on the PACT recommendations (Stewart 1992). Specifically, there are little objective data available to assess the progress made in the development and use of multi-item scales. It is the purpose of this study, therefore, to perform an inventory and evaluation of advertising-related scales. Insights from this analysis lead to specific observations regarding the state of the art in advertising research as well as recommendations for future scale research.

2. Background

In an examination of the primary thrust of all articles published in the Journal of Marketing Research between 1980 and 1986, Malhotra (1988) found that the second largest single topical category (12.5%) consisted of advertising-related studies. Moreover, Hensel and Bruner (1992) recently reported that approximately 12% of all the multi-item scales published in six leading marketing journals between 1980 and 1990 were ad-related. Yet, when Yale and Gilly (1988) examined the content of ad-related articles published in a similar set of marketing journals between 1976 and 1985, they found that only 8% of these articles directly focused on methodological issues. Furthermore, in a twenty year appraisal of Journal of Advertising articles, Muncy (1991) reported that only 5% of all articles were directly concerned with methodological and measurement issues, with fewer than one percent of the articles actually focusing on scale.

Readers with comments or questions are encouraged to contact the authors via email.
While these studies provide important insights regarding the concentration of advertising research activity by topical area (i.e., content analysis), they do not address the important issues of scale development and use within theses various construct areas. Other analyses of advertising research have likewise ignored these methodological issues, focusing instead upon such topics as quantifiable publication productivity (Barry 1990), types of research being conducted (Zaltman and Moorman 1989), and sources of new ideas in advertising (Russell and Martin 1976). An impressive study of the predictive validity of some copy testing measures has been made (Haley and Balinger 1991) but it provided little information about the psychometric qualities of multi-item scales.

Given the frequent use of scales in advertising studies, it is surprising to discover that no research has examined in detail characteristics of the measures as a group. In fact, for the field of marketing in general there have been only a few meta-analyses of scale usage (Churchill and Peter 1984; Peter 1979, 1981; Peter and Churchill 1986; Rentz 1988). As important as these studies were, there was no analysis of what constructs the scales measured. Instead, such studies have focused on evaluation of scale reliability (Churchill 1979; Peter 1979; Peter and Churchill 1986), construct validity (Churchill 1979; Peter 1981), unidimensionality (Gerbing and Anderson 1988), and generalizability (Rentz 1988).

Hence, previous meta-analyses have focused either on research concentration by advertising topic or on general characteristics of all marketing-related scales. This study serves to bridge the gap by evaluating the psychometric characteristics of ad-related multi-item scales which have been grouped by specific advertising construct categories. Specifically, the purpose of this study is to answer the following three questions: how much growth has occurred in ad scale usage over time; is scale quality improving; and, what advertising-related constructs have been measured using multi-item scales. Additionally, this study builds upon the work of Yale and Gilly (1988) as well as Muncy (1991) by providing additional indications of the ex post facto priorities of editors, reviewers, and researchers.

3. Methodology

In order to conduct this study a database was required that provided detailed information on advertising scales from a variety of sources for a period of years. However, the effort involved in performing such a census from scratch would have been prohibitively time-consuming. Fortunately, a reasonably adequate alternative was available that made the task much simpler. Specifically, a list of such scales has been recently compiled by Bruner and Hensel (1992). Their inventory was based upon a review of all research studies published in six major marketing journals for the ten year period from 1980 to 1990 that used multi-item scales.

Since information in the book was in prose form, the present study had to develop a numerical database, referring to the original articles when necessary and adding information not included in the book. Ninety-six uses of scales were found and grouped into five major categories. In order to develop the construct categories, several sources were consulted. First, a number of advertising texts were scrutinized in order to identify potential construct categories. Next, topical areas developed by Gilly and Wind (1988) as well as those derived by the Journal of Advertising for their twenty year index, which were subsequently assessed by Muncy (1991), were evaluated. This procedure resulted in the use of five major construct categories to group the scaled constructs as shown in the Table 1.

The items composing the individual scales were examined in great detail by the principal investigator. Based upon semantic similarity of items, twenty-six separate "constructs" or subcategories emerged. Then, four additional judges trained at the doctoral level were provided with a list of the twenty-six constructs as well as a compilation of all scales and their respective items. The judges were instructed to assign each scale to a construct. There was very little disagreement among the judges' assignments; only those scales composed of items which appeared to measure more than one construct presented a problem. The overall level of inter-coder reliability, calculated using the method prescribed by Perreault and Leigh (1989), was extremely high (97%).
Table 1 Categories of Advertising Constructs*

I. Advertising and Parental Influence

Parent/child interactions regarding advertising, parental concerns about advertising, parental influence over child regarding advertising effects.

II. Advertising-Related Attitudes

Attitudes and affect, attitudes toward advertising, attitude toward the ad, attitude toward the advertiser.

III. Advertising and Information

Informational content, information utility, information usage, adequacy of information provided, information processing.

IV. Advertising Effects

Recall, recognition, emotional and other types of responses, perception, awareness, persuasion, wearout, attention.

V. Advertising Practice/Management

Advertising management issues (e.g., budget, reach, sales response, promotions), media topics, strategy, public relations, professionalism, creativity, advertising agency concerns.


One point of clarification is necessary here. The majority of the articles reporting scale usage labeled the scale according to the construct area it purportedly measured. However, rather than relying on the authors' statements of what their scales were intended to study, assignment to constructs was based on examination of scale content vis-à-vis the individual scale items. This issue will be addressed further in the discussion section of this paper.

4. Findings

4.1 General Analysis

General characteristics of the database are presented in Table 2. Of the ninety-six scales composing the database, the greatest number were found in the Journal of Advertising (38.5%) and, surprisingly, the fewest were in the Journal of Advertising Research (5.2%). The format of measurement was almost equally split between semantic differential and Likert-type. Most of the scales were quite short with half having four or fewer items. It was also evident that five (37.4%) and seven point (34.1%) response scales predominated. As a group, these ad-related scales had a very respectable mean reliability (internal consistency) of 0.83. Only 11.5% of the scales reviewed reported alphas of less than 0.70, a typically used level of minimally acceptable scale internal consistency (Nunnally 1978, pp. 245, 246).

As apparent in Table 2, there is a very distinctive trend regarding scale usage across the decade of the eighties. While reported use of multi-item scales in advertising research was very light during the first half of the decade, scale use dramatically increased during the second half of the decade, with about 85% of the total decade's usage being reported from 1985 through 1989. This could in part be explained as a reaction to the PACT report which strongly advocated the routine use of multiple measures to assess the performance of ads.

Another scale characteristic worthy of note involves scale origins. A slight majority of the scales with known origins were borrowed from previous studies rather than being developed within the study in which they were first reported. However, these observations are tempered by the fact that nearly a third of all scales reviewed had no clear indication of scale origin.
Table 2 General Characteristics Of Advertising Scales

<table>
<thead>
<tr>
<th>Journal</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal of Advertising</td>
<td>37</td>
<td>38.5</td>
</tr>
<tr>
<td>Journal of Consumer Research</td>
<td>23</td>
<td>24.0</td>
</tr>
<tr>
<td>Journal of Marketing Research</td>
<td>15</td>
<td>15.6</td>
</tr>
<tr>
<td>Journal of the Academy of Marketing Science</td>
<td>9</td>
<td>9.2</td>
</tr>
<tr>
<td>Journal of Marketing</td>
<td>8</td>
<td>8.3</td>
</tr>
<tr>
<td>Journal of Advertising Research</td>
<td>5</td>
<td>5.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale Type</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likert-type</td>
<td>48</td>
<td>50.0</td>
</tr>
<tr>
<td>Semantic-Differential</td>
<td>47</td>
<td>49.0</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year Published</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980-1984</td>
<td>14</td>
<td>14.6</td>
</tr>
<tr>
<td>1985-1989</td>
<td>82</td>
<td>85.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 - 4</td>
<td>49</td>
<td>51.0</td>
</tr>
<tr>
<td>5 - 10</td>
<td>33</td>
<td>34.4</td>
</tr>
<tr>
<td>11 - 32</td>
<td>14</td>
<td>14.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of Points</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 - 4</td>
<td>4</td>
<td>4.4</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>37.4</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>8.8</td>
</tr>
<tr>
<td>7</td>
<td>31</td>
<td>15.4</td>
</tr>
<tr>
<td>unknown*</td>
<td>5</td>
<td>----</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Origin</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>28</td>
<td>43.0</td>
</tr>
<tr>
<td>Borrowed</td>
<td>37</td>
<td>57.0</td>
</tr>
<tr>
<td>unknown*</td>
<td>31</td>
<td>----</td>
</tr>
</tbody>
</table>

*The data for unknowns are included in the frequencies but are not calculated as part of the percentages.

A final item of interest involving the database as a whole is that the year of publication was significantly related to the reliability of the scales employed during that period of time (R=0.35, p<.001). Another way to view the same relationship is by noting that the mean reliability of scales published in the second half of the decade (0.84) was significantly higher than the mean for the first half of the decade (0.74). Clearly, significant strides were made in enhancing the internal consistency of advertising scales as a group.

5. Analysis By Category

The five major categories, the twenty-six subcategories, and some of their scale characteristics are presented in Table 3. There are two characteristics in particular which require closer examination. First, there are large differences in the frequency with which the constructs have been measured. Half of the construct categories have been measured only once or twice, while, at the opposite extreme, general measures of attitudes have drawn the lion's share of attention. Not surprisingly, measures of Attitude Toward The Ad have by far had the most usage, representing over 62% of all scale uses in the database. Even more specifically, general evaluative measures of Attitude Toward The Ad, such as that used by Mitchell and Olson (1981), represented nearly a quarter of all scale uses reviewed.
Secondly, there is substantial variation in both mean scale reliability (i.e., internal consistency as measured by Cronbach's alpha) as well as in the range of scale reliabilities within categories. While the majority of constructs have respectable reliabilities associated with them, some constructs appear to be inadequately measured. As evidenced by its poor internal consistency (.43), the Distracting dimension of Attitude Toward The Ad is arguably the construct most deserving of additional developmental attention. Similarly, with an alpha of .50, the measure of the adequacy of information provided by advertising for making purchase decisions seems in need of much improvement.

It should also be noted that only one scale related to the Advertising Management group was found. Apparently, the thinking of ad managers was not studied much using psychometric measures. A similar observation can be made regarding the barely measured and yet increasingly important constructs related to Attitude: Parental Influence. There seems to be an important opportunity here for scale development and the insights such tools can bring to these topics.
6. Discussion

Although the process of developing a database was simplified by use of information provided in the Bruner and Hensel (1992) compilation, it was not without its drawbacks. For example, there are other sources of ad scales that were not included in the book such as the journals closely linked with the journalism and communication fields. There are also non-publicized sources of scales such as the companies and agencies that use them in the industry on a daily basis. Further, it is not known to what degree multi-item scales were used to perform ad research in non-English speaking countries. Therefore, it may be safest to view these data described here as representing a portion of all ad scales used in the 1980s and skewed toward research conducted by marketing professors in English-speaking North America.

Based on their findings, the PACT coalition strongly advised the routine use of multi-item measures in assessing the performance of ads. It is evident from the analysis of the data presented here that advertising research moved toward this goal during the decade of the 1980s. PACT also stipulated that the reliability and validity of measures should be empirically demonstrated. Progress was made in establishing the internal consistency of scales but such growth was not clear for the other types of reliability nor the several forms of validity. It was not expected that sophisticated examination of validity would be common; however, it was surprising that testing of such fundamental qualities as content validity and unidimensionality were not routinely provided. Certainly, the examination of validity is complicated by the inconsistent use of terminology and a lack of agreement on the methods of assessment. So, little evidence exists which suggests that researchers have attempted to demonstrate (let alone improve) the validity of their measures.

In reviewing the collective investigations of advertising constructs through the use of multi-item scales, one is struck by the overwhelming emphasis that has been placed on Attitude Toward The Ad and how little other ad-related constructs have been examined using multi-item scales. Using the list of topics developed by Yale and Gilly (1988) and Muncy (1991) for their content analyses, it appears that published advertising research involving several frequently investigated topical areas have not incorporated the use of multi-item scales in their studies. Probably the most obvious gaps are in social issues and advertising practice research. Although articles focusing on these issues composed 56% of the total articles in the domain studied by Yale and Gilly (1988), presently available scales are sorely inadequate for their continued investigation.

Moreover, in the early 1980s, a task force of the Marketing Science Institute (MSI) formulated three categories of top priority advertising research topics (Schmalensee 1984): (1) theory of individual consumer response to advertising, (2) measures of individual consumer response to advertising, and (3) advertising operating concerns. Of these, the vast majority of the research published (employing multi-item scales) in the investigated journals addresses only the first topic. Very little attention has been paid to either of the second two priority questions.

A related observation is that many researchers have attempted to compare previous findings based upon the apparent equivalency of the measures studied even though the content of those measures is quite different. For example, studies by Holmes and Crocker (1987), MacKenzie and Lutz (1989), and Stout and Burda (1989) all reported measuring something they referred to as Attitude Toward The Ad. Yet, not only do their three scales have completely different bipolar adjectives, but they appear to measure different aspects of attitudes. Hence, any meta-analysis of effects related to Attitude Toward The Ad would need to take into account that varying results could be explained by varying constructs being measured even though all have been referred to as the same thing.

When selecting scales researchers must be very careful to distinguish between measuring a person's perception of an ad and the emotions evoked by an ad. For example, the Ad-Generated Emotions scales in Table 3 were intended to measure a person's self-reported emotional reactions to ad. In contrast, the scales measuring the constructs under Attitude Toward The Ad measure a person's evaluations of various aspects of the ad itself. These are different though related constructs which should be carefully distinguished by researchers (Burke and Edell 1989; Edell and Burke 1987).

Based upon these and other findings, we repeat PACT's call that there must be greater concern for construct measurement. As initial steps, journals should consider instituting the following desirable practices: a moratorium on the publication of articles using "quick and dirty" scales when satisfactory measures are already available in the literature; provision of sufficient evidence in manuscripts to indicate that scales are reliable and valid; and, inclusion of scale items in
articles so that others may judge their face validity and potentially use them in their own research. These issues stand a greater chance of being addressed by both authors and reviewers if submission procedures such as those implemented by the *Journal of Marketing* (Kinnear 1992) and the *Journal of Consumer Research* (Monroe 1991) are adopted by more advertising-related journals.

7. Suggestions for Future Research

As a means of showing that the professional organizations related to the field are serious about measurement issues, we recommend a journal be started which has measurement and methodological issues as key topics. If that is too difficult to accomplish anytime soon then the next best thing might be for the established journals in the field to have special sections devoted to measurement issues such as was instituted several years ago in the *Journal of Personal Selling and Sales Management*.

In terms of future research, periodic assessments such as conducted here would appear to be desirable in order to determine if progress is being made. Although assessment can be done using a set of academic journals as done here it would also be worthwhile to determine what measures are really being used in industry versus academia. For example, it could be that lots of “good” scales are being generated in academia but they are unknown or unused by practitioners. On the other hand, it could be that lots of measures are being generated and used in industry but they are of unknown quality because of the lack of time or interest to conduct such laborious validation studies. Beyond that, there is always the opportunity to step outside this country and see what the state of the art is in other parts of the world both in academia as well as industry.

With the rise in e-commerce over the last few years and now the birth of m-commerce, some scales will be carried over from there usage in more traditional contexts while others will be adapted. Of course, many will be developed anew. Hopefully, with the greater ability to conduct literature reviews and know what measures are available before beginning a study, we can avoid the tendency to “reinvent the wheel.” Yet, early indications are not encouraging. For example, one of these new constructs that appears to have a major impact on ads seen at a website is attitude-toward-the-website (Bruner and Kumar 2000). Already, numerous attempts to capture this construct have been made but it is unknown to what extent one measure is “better” than the others or if, in fact, they may be measuring different things (e.g., Burns 2000; Chen and Wells 1999; Thistlethwaite and Bauerly 2000; Wu 1999).

Future advertising researchers will continue to be faced with decisions regarding scale construction, modification, or adoption. While a number of existing scales have been identified both here and elsewhere (Bearden et. al 1993; Bruner and Hensel 1992), it is evident that many of them require further development before they should be employed in future studies. Adoption and/or modification of existing scales allows the researcher to make parsimonious use of his or her time while further refining such scales. Moreover, as noted by Shaw and Wright (1967), quality of scales and standardization of research often is worse when the researcher devises new scales than when the researcher builds upon existing scales. Nevertheless, investigators should carefully evaluate existing scales, and judiciously adopt them only when they measure the specific construct of interest in a valid and reliable way.

8. Summary and Conclusions

A decade has passed since release of the PACT statement. This study offers an initial examination of the progress made in response to the PACT recommendations. Answers to three key questions have been provided: usage of multi-item scales appears to have grown dramatically over time; the quality of those scales as measured by internal consistency has improved; and, scale usage has clustered heavily around the Attitude Toward The Ad construct area.

What remains is to determine how changes in the usage of multi-item scales has progressed in industry as opposed to academic research as primarily true here. The cooperation of the agencies and companies conducting such research will be necessary rather than relying on journals for data. Further, periodic assessment of scale usage would seem to be a desirable goal in order to chart the progress being made by the field. It is hoped that subsequent studies will be more successful than this one in addressing issues involving validity. If these matters are effectively dealt with by academic and industry researchers in the next few years then we can reasonably expect the decade of the 1990s to end with
the concerns identified by the PACT coalition to have been successfully addressed.

Footnotes

1. Greater detail regarding criteria for inclusion in the book can be found in Bruner and Hensel (1992, pp. ix, x).
2. A fuller examination of scale reliability and validity was part of the original intent of this study. However, in the process of developing and refining content codes, so few instances of validity testing occurred that the focus shifted towards information that was routinely reported, e.g., internal consistency. Among the forms of psychometric quality that were rarely if ever tested were stability (test-retest), alternate forms reliability, convergent validity, and discriminant validity. Although not routine, unidimensionality was a little more commonly addressed but only via exploratory factor analysis.

References


