
Standardization & Justification: Do A_{ad} Scales Measure Up?

Gordon C. Bruner II

Several previous reviews of A_{ad} have noted that progress in understanding advertising's effects is hampered by the lack of a shared view of attitude structure. This study builds upon those reviews by focusing on two additional problems: measure standardization and justification. Examination of a decade-and-a-half's worth of research found 75 multi-item measures of A_{ad} involving 53 different semantic differentials. A cluster analysis indicated that while three different groups of scales could be identified there was a low amount of standardization within clusters. Further, it was rare for authors to provide justification for their scales. Practical recommendations are provided about how these shortcomings can be addressed.

Articles written by Mitchell and Olson (1981) as well as Shimp (1981) are generally credited with introducing and suggesting the importance of attitude-toward-the-ad (A_{ad}) as a mediator of advertising's effects on several variables of relevance to marketers such as brand attitudes and purchase intentions. Following Mitchell and Olson (1981), most researchers operationalize A_{ad} using semantic differential scales (Muehling and McCann 1993). As popular as this approach has been, however, there is little consistency in the items used.

The purpose of this paper is to examine some specific issues that affect the way A_{ad} is measured and reported. Admittedly, questions about the definition and operationalization of A_{ad} have been raised in the past, but no detailed examination of the extent of the problem has been published. In particular, this paper focuses on the twin issues of *standardization* and *justification*, that is, consistent use of the same measure for the same construct and the provision of adequate reasoning and measure validation when alternative measures are used. Not only does this paper attempt to gauge the extent to which a lack of standardization and justification have affected A_{ad} studies but it will also suggest some ways this neglect can be addressed.

Background

One of the most thorough meta-analyses published in recent years in the field of advertising focused on

studies of A_{ad} (Brown and Stayman 1992). Because significant variation across studies was found, a number of methodological variables were identified as potential moderators of the relationships. A methodological issue not given in-depth consideration was the diversity of conceptualization and operationalization of A_{ad} .

The possibility exists that the variety of instruments used over the years have not had as much in common as their names might suggest (Percy and Rossiter 1992). Indeed, concern has been expressed over the last decade that researchers are not carefully specifying the domain of the A_{ad} construct and are not validating their measures (Allen and Madden 1989; Burton and Lichtenstein 1988; Wright 1986). It appears that a *mélange* of scales have been produced over time with little observable similarity except that they are purported to measure something related to A_{ad} (Bruner and Hensel 1996, pp. 809-825; Muehling and McCann 1993; Wiles and Cornwell 1991).

Why has this situation developed? The primary and most obvious reason is that there is no one accepted theory of attitude structure (Eagly and Chaiken 1993). Specifically, A_{ad} has either been viewed as multi-dimensional or as unidimensional. In A_{ad} studies the unidimensional view has led to the use of one general evaluative measure (e.g., Mitchell and Olson 1981), whereas the multi-dimensional view has required the use of two or more measures (e.g., Baker and Churchill 1977). The two most common components are referred to as

Gordon C. Bruner II (Ph.D., University of North Texas) is Associate Professor, Department of Marketing, Southern Illinois University, Carbondale, Illinois.

cognitive (thoughts regarding individual characteristics of an object) and *affective* (overall feelings about the object). (For a more detailed review of the competing conceptualizations see Muehling and McCann 1993.) Those who have critically reviewed A_{ad} literature in recent years have concluded that no well-accepted conceptualization of the construct has emerged and it will be difficult to draw more insightful conclusions about A_{ad} 's antecedents and consequences until the definitional issues are worked out (Muehling and McCann 1993; Percy and Rossiter 1992).

Despite this observation, the position taken here is that scale diversity is not explained solely in terms of adoption of different theories of attitude structure by researchers. Even within a single conceptualization of A_{ad} there has been a surprising lack of common scale usage (standardization). For example, both Yi (1993) and Homer (1990) appeared to embrace the single component model of A_{ad} and thereby employed one scale to measure it. However, the scales they used did not have any items in common. Likewise, both Petroschius and Crocker (1989), as well as Zinkhan and Zinkhan (1985), held the dual component view of A_{ad} , but, of the fifteen different items that composed their scales, only two were held in common.

In addition, justifying the use of a particular measure does not appear to be a priority. It is not unusual for many apparently acceptable alternative measures to be ignored in favor of new scales being introduced with little, if any, published reasoning or validation. This study focuses on these twin issues, standardization and justification, and the degree to which they appear in A_{ad} research.

Standardization

Standardization is used here to refer to *the extent to which the same rules have been used to assign numbers to a construct*. This perspective is not new. It can be found in Nunnally (1978, pp. 3-9) but was more explicitly described in the newer edition (Nunnally and Bernstein 1994):

A measure is standardized to the extent that (1) its rules are clear, (2) it is practical to apply, (3) it does not demand great skill of administrators beyond that necessary for their initial training, and (4) its results do not depend upon the specific administrator. The basic point about standardization is that users of a given instrument should obtain similar results (p. 4).

They go on to state that a fundamental principle of science is that any observation made by one researcher should be independently verifiable by other researchers and this "principle is violated if scientists can disagree about the measure" (Nunnally and Bernstein 1994, p. 6). If alternative rules are used and the relationship between the resulting measures is known then that might be acceptable as well. But, the point is that standardization is lacking when a variety of different rules are used by researchers to produce alternative measures of supposedly the same construct and the relationships among the measures are unknown.

With specific reference to rating scales, there are several reasons for standardizing usage. First, some degree of standardization is necessary in order to validate a scale because validation is best viewed as a process; it is not likely to occur in one study but requires the methodical testing of a measure in multiple studies to produce a knowledge base of its psychometric properties (e.g., Cronbach 1971; Peter 1981). Similarly, no single use of a scale is likely to examine all of the facets of generalization that should be addressed by any measure for which wide application is desired (Rentz 1987; Finn and Kayandé 1997). Third, a proper "final" step in scale construction is development of norms (Churchill 1979). This is helpful because scale scores are best interpreted in light of normative data regarding a measure's use with different groups and in a variety of situations (Furse and Stewart 1982). Thus, multiple studies across conditions and populations should be conducted over time utilizing the same scale in order to achieve the important psychometric goals of measure validation, generalizability, and norm development.

Beyond these general reasons for the use of standardization a more immediate concern with regard to A_{ad} is that diversity of operationalization can affect the conclusions that are drawn regarding the presence of significant relationships with other variables of interest. For example, in several studies where multiple measures of A_{ad} were used, the conclusions drawn about the significant relationships of one measure were different from those based on the other measure (Burton and Lichtenstein 1988; Burton and Zinkhan 1987; Okechuku and Wang 1988; Olney, Holbrook, and Batra 1991; Petroschius and Crocker 1989). In particular, the study by Miniard, Bhatla, and Rose (1990) illustrates that the role of one attitudinal component can be dis-

tinct from another component, while a global measure can lead to still different conclusions.

Lack of standardization has been acknowledged as a problem in our parent disciplines for some time. For the period of 1954 to 1965 in sociology journals, 72% of indices and scales were found to have been used just once (Bonjean, Hill, and McLemore 1967). In psychology for the period of 1960 to 1969, 63% were used just one time (Chun, Barnowe, Cobb, and French 1974). No similar study of usage frequency has been published about marketing measures as a whole but it has been reported that 79% of the scales published in consumer studies during in the 1980s were used just once (King and Bruner 1993). In cases such as these where scales are developed, used once, and then rarely if ever used again it suggests the research process has been "discontinuous, fragmented, and wasteful" (Chun, Cobb, and French 1975, p. x).

Justification

As used here, justification refers to the task of *providing adequate reasoning for the development of a new measure as well as evidence of its validity*. Certainly, some constructs are more important to a study than others and space is limited. But, the more central a construct is to the heart of the study then the more stress should be placed upon justification.

The problem comes when standardization and justification are not considered to be high priorities when study methodology is planned. For example, one could argue that standardization and justification are unnecessary if a new scale's items have been drawn from the same semantic domain as alternative measures because they all can be assumed to have a similar amount of common core. However, the strength of this position depends upon the extent to which items for the alternative scales have been properly sampled from the same semantic domain and evidence of their congruity is published. In such a case the burden of justification is on the authors themselves because the production of new measures when established alternatives are available demands explanation (Varadarajan 1996). Indeed, "researchers should have good reasons for proposing additional new measures given the many available . . . and those publishing should be required to supply their rationale" (Churchill 1979, p. 67).

Providing evidence of psychometric quality is most called for when a new scale is used but it is also important for measures that are modified in some way. In psychology, a primary standard of measure use (as opposed to a conditional standard) is that it is incumbent upon subsequent users of a measure who modify it in some way to revalidate it or at least explain why additional validation is unnecessary (AERA, APA, and NCME 1985, p. 41).

Even when an established scale with known psychometric properties is used some limited justification would be helpful. It may be as simple as citing the relevant sources where evidence of the scale's validity can be found. The confusion comes when the source of the scale is not provided and evidence of validity is lacking as well. Readers must guess whether the scale is new, adapted, or borrowed as well as the degree to which it is a valid measure. This guessing and uncertainty decreases confidence in the findings associated with the measure. Although we have tolerated such practices in marketing, one expert with fifty years of experience examining measures in psychology has urged researchers to be suspicious when undocumented scales are encountered, despite who authored the study or where it was published (Buros 1975, p. xvii).

Goals of Current Study

The foregoing discussion has offered a couple of additional reasons for the problems associated with A_{ad} measurement beyond the one previously identified by others (disagreement about attitude structure). It is the aim of this study to estimate the extent to which standardization and justification have characterized A_{ad} studies by conducting a detailed analysis of the scales themselves. Two primary research questions and related issues will be addressed via an examination of past studies of A_{ad} .

The first primary research question involves standardization: to what extent has there been high diversity of item content between scales based upon different conceptualizations and high commonality of item content within scales from the same conceptualization? Given the two major conceptualizations of A_{ad} , unidimensional and multidimensional (Muehling and McCann 1993), it is expected that there is substantial variance in measurement between scales linked to the different views. However, it is also expected that there should be low diversity (high standardization) for scales

that share the same conceptual background. Given this, three primary and distinct clusters of scales are anticipated: a group measuring A_{ad} using the single-component approach and two groups (e.g., affective and cognitive) related to the dual component view of A_{ad} .

The second research question focuses on justification: to what degree have articles provided reasoning and support for the particular A_{ad} scales used, especially the new and modified measures? Due to the stress placed on psychometric quality in several key marketing-related articles just prior to the flood of A_{ad} studies in the 1980s (e.g., Churchill 1979; Jacoby 1978; Peter 1979), a high premium should have been placed upon justification. If justification of A_{ad} scales is important to authors, editors, and reviewers then it is expected that information about a scale's source, selection, and validity would be provided to readers.

Methodology

Measures included in the review were identified from examination of seven journals over a fifteen year period (1980-1994): *Journal of Advertising*, *Journal of Advertising Research*, *Journal of Current Issues and Research in Advertising*, *Journal of the Academy of Marketing Science*, *Journal of Consumer Research*, *Journal of Marketing*, and *Journal of Marketing Research*. While A_{ad} measures can be found in other journals, conference proceedings, and time periods, the domain examined here was considered to be more than adequate to illustrate the problems of standardization and justification. In fact, such a cumbersome array of measures was found that a further means of focusing the review was necessary. Specifically, only those single- and dual-component scales using at least three sets of bi-polar adjectives and for which the items were known were selected for greater scrutiny. This appears to be the overwhelming favorite method for measuring the construct (Bruner and Hensel 1996, pp. 820-825). However, this limitation meant that some shorter scales (e.g., Stout and Burda 1989), Likert-type scales (e.g., Zinkhan, Locander, and Leigh 1986), and measures of third components from tripartite models were left out (e.g., the *interestingness* component used by Olney, Holbrook, and Batra [1991] and the *conative* component used by Okechuku and Wang [1988] as well as others). Including these other scales would only have increased the variance found between measures.

To address research question 1, the unique bi-polar adjectives were numbered (Appendix 1) and the set of items belonging to each multi-item scale were dummy coded to indicate their presence or absence (Appendix 2). Cluster analysis was employed as an objective means of helping to determine which scales (sets of items) had enough common content so as to be considered the "same" measure and which ones were so diverse that they should be considered different measures. The FASTCLUS routine within SAS was used because it is a well accepted, nonhierarchical approach and is recognized as an effective means of identifying outliers (Hair, Anderson, and Tatham 1987, p. 332; SAS Institute 1988, p. 494).

The recommended multi-stage process for determining the optimal cluster solution was followed (SAS Institute 1988, pp. 502). Briefly, this amounted to performing a preliminary FASTCLUS with a large number of clusters, noting the number of clusters with very low membership, and then deleting the outliers from the development of cluster seeds. Determining the optimal cluster solution was primarily guided by noting the maximum values of the Pseudo F, a statistic found to outperform other criteria (Milligan and Cooper 1985) and one with previous use in marketing research (e.g., Bowen 1990). It has been recently suggested that two additional statistics be examined as well (Sharma 1996, pp. 194-211). *Within Standard Deviation* is the root-mean-square total sample standard deviation pooled across all of the clusters where smaller values indicate greater homogeneity within clusters. *R-Squared* is the ratio of sum-of-squares (between) divided by sum-of-squares (total) and is interpreted such that higher values (approaching 1) indicate greater homogeneity of the clusters.

The second research question was investigated by noting for each use of an A_{ad} scale what sort of justification was provided for the specific measure used. This was difficult to assess given that no one piece of information provides objective evidence of justification or lack thereof. The position taken was that justification is proper for all scales and a minimal form occurs when authors borrow a previously tested scale from published research and cite the source. First it was noted what the authors stated about the source of a scale. Given that some scales were likely to have been taken from previous research, though not mentioned in the published version of the article, several pieces of additional information were collected and considered. Having

all of the items for these scales as well as the cites made by the individual authors allowed a detailed comparison of item sets and facilitated an independent check of a scale's source. (See Appendix 3 for more detail regarding the determination of scale source.)

Another form of justification is called for when new or greatly modified scales are produced. As evaluated here, the inadequacy of previous measures should be explained and initial evidence of the scale's validity should be provided. If, instead, the scale's content is simply listed with no evidence of validity, it was considered to illustrate a lack of justification.

Findings

Scales from 62 articles utilizing 53 different bipolar adjectives were identified in the domain of review. Due to the fact that some studies used more than one measure in order to capture multiple attitude components, a total of 75 scales composed the database.

Standardization

Statistics of the FASTCLUS were used to evaluate research question 1 and are given in Table 1. Initial unrestricted runs of the cluster analysis indicated that there were only seven clusters that had two or more members. Given that, cluster solutions of two through seven were more closely scrutinized. More support was found for the three cluster solution than the others. Not only did it have the highest Pseudo F, but the largest drop in the RMSSTD occurred when moving to the three cluster solution from the two. The three cluster solution was also able to accurately separate into different groups the pairs of scales as used by Petroschius and Crocker (1989) and others that had few if any items in common, i.e., those that were referred to as measures of the cognitive component were consistently grouped together and those referred to as measures of the affective component were assigned to a different cluster.

Even though the three cluster solution was considered to have more support for it than any other grouping, it resulted in a low R^2 (.241). This highlights the fact while there was enough structure in the data to identify three groups of scales, there was low homogeneity of content within the clusters. Given this, interpretation of the groups is not simple though some observations are possible.

The bulk of the single-component scales grouped in cluster 1. With 51 members, cluster 1 is also the largest group. Based upon the names given to these scales by most of their users, this cluster may be best described as representing global evaluative measures of A_{ad} . Only a few of the single-component scales were in clusters 2 and 3. This supports the expectation that single-component scales have more in common with other single-component scales than with those measures that stem from a multi-component view.

Clusters 2 and 3 each had twelve members. While not in total agreement, names given to these scales by their authors indicate that cluster 2 is related to an affective dimension of A_{ad} . In contrast, the scale names indicate that cluster 3 relates to a cognitive dimension of A_{ad} . Cluster statistics show that clusters 2 and 3 are the most distinct groupings of the three clusters. This is reasonable since the researchers who used a pair of scales to tap into different subconstructs rather than one global measure deliberately employed different sets of items.

Justification

Part of addressing research question 2 was determining the extent to which researchers justified their scale selection in some way. To make this judgment required knowing the source of scales. As can be seen in Table 2, very few of the articles indicated clearly whether the scales had been borrowed from previous research, were modifications of previously used measures, or were original.

An independent examination of cited articles indicated that 33% (25) of the scales were borrowed intact from previously reported studies despite whether such studies were identified by their users. About 37% (28) seemed to be modifications of previously used A_{ad} measures, while 25% (19) were judged to be original. The source of the remaining scales (3) could not be determined with a sufficient degree of certainty. Only 12% of the scales were accompanied by even the slightest explanation for their selection and use. Admittedly, some authors indicated they used *items* that had been used before but the focus here was to note those instances where there was reasoning provided for a particular *set of items* (the scale as a whole).

Research question 2 was also examined by identifying what portion of scales were accompanied by information about their reliability, dimensionality, and validity. Reliability was reported for most scales (95%), and in the majority of those cases (79%), the

Table 1
Results of Cluster Analyses

Number of Clusters	Pseudo F	R ²	RMSSTD
2	8.57	.105	.335
3	11.44	.241	.217
4	9.76	.292	.212
5	10.39	.372	.200
6	8.81	.390	.199
7	9.15	.447	.191

levels were above .80, the developing standard for scales used in basic research (Peterson 1994). In contrast to the high incidence of reporting of reliability, only 32% of the scales provided evidence that dimensionality had been tested. In each of these cases the evidence indicated (or the authors stated) that the scales were unidimensional.

For each study it was also noted whether evidence was provided of a scale's convergent and/or discriminant validity. While a certain pattern of relationships reported in a study could be construed *post hoc* by others as providing an indication of a scale's nomological validity, it was not considered here to be evidence unless the authors themselves stated it as such. The findings indicate that the validity of A_{ad} scales has been rarely examined. Only the authors of six studies (representing eight scales) reported evidence bearing directly on validity.

Discussion and Recommendations

Examination of the two research questions led to the following conclusions. First, there is a lack of scale standardization in measuring A_{ad} not only because different conceptualizations of attitude structure are being used but because, even within similar conceptualizations, new and modified scales are routinely produced. Second, standardization is further hampered because of the lack of scale selection justification; that is, when new or modified scales are introduced there has not been sufficient reasoning provided for their use nor adequate evidence provided of their validity.

The results of this review show that researchers have heeded the call to make sure their measures are reliable. However, evidence of unidimensionality and validity are much less frequently provided.

Even those scales that have provided some evidence of validity lack proper justification when numerous previously developed scales were ignored in favor of constructing a new one and no clear reasoning for the new scale was provided.

Although much could be made out of the lack of clarity and inaccurate citations, the bottom line is that very few authors gave much explanation of their scale choice. Early workers in the field (e.g., Mitchell and Olson 1981) can not be faulted because they produced new scales at a time when few if any alternatives existed and standards of scale construction were less widely accepted. However, as time went on it is not clear why the scales being used were much more likely to be new or modified rather than borrowed intact from previous research. Maybe researchers produced new scales due to some uniqueness of the ads under study, or they considered previous measures to lack validity, or because they simply wanted a proprietary measure. But, in all cases, if authors' considered it important to modify scales or create new ones, it is unfortunate that it was not considered just as important to provide reasons for those actions.

Again, it is critical to emphasize that the review and conclusions here are greatly affected by the information available in the published versions of articles. The issue is not just whether authors have tested the validity of a scale but whether it was reported in the published version of the article. No doubt space limitations and other issues have affected the extent of what is reported. The point is that provision of validity information puts other researchers on notice about the quality of a measure and the burden is on them to take it into account when they select a measure to use. In the absence of such information being reported, read-

Table 2
Justification of A_{ad} Scales

<i>Authors (date)</i>	<i>Scale Name</i>	<i>Scale Article</i>	<i>Source¹ Check</i>	<i>Evidence of Validation²</i>	<i>Reasoning Provided³</i>
Andrews et al. (1992)	A _{ad}	M	M	R	NO
Boles & Burton (1992)	A _{ad}	B	B	R	NO
Buchholz & Smith (1991)	A _{ad}	B?	O	R	NO
Burton & Lichtenstein (1988)	cognitive	O	O	R*, D, V	NO
Burton & Lichtenstein (1988)	affective	O	O	R, D, V	NO
Chattopadhyay & Basu (1990)	A _{ad}	?	M	R	NO
Chattopadhyay & Nedungadi (1990)	A _{ad}	?	M	R, D	NO
Cox & Cox (1988)	Ad eval.	?	B	R	NO
Cox & Locander (1987)	Ad eval.	M?	M	R	NO
Darley & Smith (1993)	A _{ad}	M?	M	R*, D, V	NO
Donthu (1992)	A _{ad}	B	M	R	NO
Droge (1989)	A _{ad}	B?	M	R+	NO
Gardner (1985)	A _{ad}	B	B	R+	NO
Goodstein (1993)	A _{ad}	?	M	R	NO
Hastak & Olson (1989)	Ad eval.	?	M	R	NO
Hill (1988)	Global A _{ad}	M	M	R	NO
Hill (1988)	emotional A _{ad}	B	M	R	NO
Hill (1989)	Global A _{ad}	M	B	R	NO
Hill (1989)	emotional A _{ad}	B	B	R	NO
Holbrook & Batra (1987)	A _{ad}	?	O	R	NO
Homer (1990)	A _{ad}	?	?	R	NO
Janiszewski (1988)	Ad eval.	?	?	R, D	NO
Kamins (1990)	A _{ad}	M	O	R	NO
Kamins, Marks, & Skinner (1991)	A _{ad}	M	M	R*	NO
Kellaris, Cox, & Cox (1993)	A _{ad}	B	M	R	NO
Keller (1987)	A _{ad}	B?	M	R	NO
Keller (1991a)	A _{ad}	?	B	R	NO
Keller (1991b)	A _{ad}	?	B	R	NO
Kilbourne (1986)	cognitive	B	B	R*	SOME
Kilbourne (1986)	affective	B	B	R	SOME
Kilbourne et al. (1985)	cognitive	M	M	R*	NO
Kilbourne et al. (1985)	affective	M	M	R*	NO
Laczniak & Muehling (1993)	A _{ad}	?	M	R	NO
Lord, Lee, & Sauer (1994)	A _{ad} (claim)	M?	M	R*, D	NO
Lord, Lee, & Sauer (1994)	A _{ad} (nonclaim)	M?	B	R, D	NO
Machleit, Allen, & Madden (1993)	A _{ad}	B	B	R, V	NO
Machleit & Wilson (1988)	A _{ad}	?	O	R, D	NO
MacInnis & Park (1991)	A _{ad}	?	M	R, D	NO
MacInnis & Stayman (1993)	A _{ad}	?	M	R, D	NO
MacKenzie & Lutz (1989)	A _{ad}	?	?	R	NO
MacKenzie & Spreng (1992)	A _{ad}	?	B	R	NO
Macklin, Bruvold, & Shea (1985)	A _{ad}	?	O	R	NO
Madden, Allen, & Twible (1988)	Ad eval.	O	O	R, D, V	YES
McQuarrie & Mick (1992)	Ad liking	?	M	R	NO
Miller & Marks (1992)	A _{ad}	B?	B	R, D	NO

(Continued)

Table 2
Justification of A_{ad} Scales (continued)

Authors (date)	Scale Name	Scale Article	Source ¹ Check	Evidence of Validation ²	Reasoning Provided ³
Miniard, Bhatla, & Rose (1990)	A_{ad} (claim)	O?	M	R, D, V	YES
Miniard, Bhatla, & Rose (1990)	A_{ad} (noclaim)	O?	M	R, D, V	YES
Miniard, Bhatla, & Rose (1990)	A_{ad}	?	M	R	NO
Mitchell (1986)	A_{ad}	B	B	R	NO
Mitchell & Olson (1981)	A_{ad}	?	O	R, D	NO
Muehling (1987)	A_{ad}	O?	O	—	NO
Muehling & Laczniak (1988)	A_{ad}	M?	O	R, D, V	NO
Muehling & Laczniak (1992)	A_{ad}	M?	B	R	NO
Muehling, Laczniak & Stoltman (1991)	A_{ad}	M?	B	R, D	NO
Muehling, Stoltman, & Mishra (1990)	A_{ad}	M?	M	R, D	NO
Okechuku & Wang (1988)	cognitive	B	B	R*	NO
Okechuku & Wang (1988)	affective	B	M	R	NO
Olney, Holbrook, & Batra (1991)	utilitarian	M?	O	R, D	YES
Olney, Holbrook, & Batra (1991)	hedonism	M?	O	R, D	YES
Perrien, Dussart, & Paul (1985)	cognitive	O	O	R*	NO
Perrien, Dussart, & Paul (1985)	affective	O	O	R*	NO
Peterson, Wilson, & Brown (1992)	A_{ad}	?	O	R	NO
Petroshius & Crocker (1989)	cognitive	B	B	R*, D	NO
Petroshius & Crocker (1989)	affective	B	B	R*, D	NO
Prakash (1992)	A_{ad}	B?	B	R+	SOME
Severn, Belch, & Belch (1990)	A_{ad}	?	O	—	NO
Singh & Cole (1993)	A_{ad} (eval.)	B	M	R, D	SOME
Smith (1993)	A_{ad}	B?	B	R	NO
Steenkamp & Baumgartner (1992)	A_{ad}	B	M	R	NO
Sujan, Bettman & Baumgartner (1993)	Ad Eval.	?	B	R	NO
Tripp, Jensen, & Carlson (1994)	A_{ad}	B	B	R, D	NO
Yi (1990)	A_{ad}	?	B	R	NO
Yi (1993)	A_{ad}	?	B	R	NO
Zinkhan & Zinkhan (1985)	cognitive	?	O	—	NO
Zinkhan & Zinkhan (1985)	affective	?	O	—	NO

¹A comparison was made between what the article indicated the source of the scale was with what was determined from an independent check made in this study. (See Appendix 3.) M=modified B=borrowed O=original ?=uncertain

²If evidence was provided in support of reliability, unidimensionality, or validity it was signified with an R, D, and V, respectively. An asterisk (*) indicates that reliability was below .80 and a plus sign (+) indicates that multiple alphas were reported rather than one overall reliability coefficient, with one above .80 and one below.

³An indication of whether or not reasoning was provided for the particular scale used in a study. In particular, was justification given for the modification of an older scale or the construction of a new scale?

ers have little or no basis upon which to conclude that one scale is any better than the many others that are also reported without evidence of validity. They may further decide that it is an acceptable practice to develop their own measures and as long as they have reasonable internal consistency they will be considered satisfactory measures regardless of their congruity with past measures. Indeed, 36% (27) of the measures reviewed here were just that

way: not only do they represent unique sets of items but they lack justification (reasoning for their use as well as evidence of validity).

In a related issue, it is not clear why authors have not been more forthcoming about the sources of their scales. Maybe it is thought that if it is admitted their measures are original, it will place their work under greater scrutiny. It is also troubling that in several cases the statements made in

the articles suggested that scales were borrowed from previous studies when the independent checks made here found enough changes to suggest that they were modifications if not altogether original scales. Thus, what we mean by *original*, *borrowed*, *adapted*, and *modified* may need some standardizing as well as our measures.

An additional problem in the scales reviewed is a lack of attention to the premise underlying the semantic differential. The items should be constructed so that the poles are adjective pairs describing opposites of the semantic continuum. It is arguable whether this requirement is being met in those many cases where adjectives of the form *X/not X* have been used. The possibility exists that items of this form compromise the assumption that the midpoint of the scale is meant to be used when respondents associate the object with neither pole of the adjective pair (Dawes and Smith 1985, p. 534; Osgood, Suci, and Tannenbaum 1957, pp. 29, 83). For purposes of this study, such differences were overlooked, but the question remains whether this is yet another source of significant measurement variation.

Recommendations

A favorable though unrealistic goal for the near future would be for researchers to adopt the same theoretical framework for A_{ad} . As this study has pointed out, however, even with the dominance of one conceptualization there would probably be a great deal of measure variance if standardization and justification are not given higher priority.

As a practical first step then, the sentiments of previous reviewers are echoed (Muehling and McCann 1993; Percy and Rossiter 1992): researchers must have an understanding of the alternative views that coexist at this time and make an informed choice between them based upon the theoretical network to which they expect their study to contribute. This should be plainly stated in the published version of the paper. A related suggestion is that researchers must be much more sensitive about trying to identify and select from among the scales used previously in work upon which they are building. Surely a previously developed scale can be found that is suitable for the study's purposes and has evidence of reliability and unidimensionality if not validity, too.

The repeated use of the same scale(s) is also encouraged for purposes of validation. There is only

so much that can be done in a single study; thus, a scale should be used and evaluated in multiple studies. This review suggests that while validity testing has begun for a few of the scales, there is no evidence of a validation process. Even for the Mitchell and Olson (1981) scale, which has been used intact more than any other A_{ad} measure, there has been no known coordinated effort to validate it, thoroughly examine its generalizability, and develop norms for its use.

If A_{ad} is one of the primary constructs in a study, then more attention should be devoted to justification than it has received previously. In contrast, when A_{ad} is peripheral to the main hypotheses of a study, the matter can rather easily be dealt with via standardization: borrow an established scale used in similar research that has shown some evidence of validity and cite the source. On a practical level this should be the minimum level of justification. Thus, justification is made easier by standardization.

It is recommended that manuscript guidelines place greater emphasis on measurement justification. New or greatly modified measures should be accompanied with both the reason(s) why the available alternatives were not satisfactory as well as some initial evidence of validity. Information bearing on justification is important enough to remain in the final published version of the paper if at all possible. If it can not remain, then a footnote should indicate that such information is available by contacting the author. But, one way or another, the justification of new measures should be produced and readers should be aware of its existence.

The field would also be served if empirical comparisons of A_{ad} scales could be conducted and published. Of most interest might be studies that examine the relative abilities of various competing scales to explain variance in other variable(s) of interest (e.g., Ab, BI). Up until now the studies have examined the scales one at a time (or as complementary pairs) rather than for the purpose of simultaneously comparing their nomological validities. Assuming some clear conclusions are drawn, authors of subsequent studies can cite such critical comparisons as justification for use of a particular scale without having to go into much greater detail.

Finally, recommendations about which scale to use in measuring A_{ad} should be guided by the known psychometric qualities of the alternatives as well as which previous studies one is building upon. Yet, with so little validation work having been con-

ducted, it is difficult to recommend any one scale at this point. Some tentative recommendations can be made, however. To begin with, if one wants a global measure of the A_{ad} construct then the Mitchell and Olson (1981) scale would appear to be a reasonable choice. Not only is it very well accepted (having been used much more than any other single scale) but there is even some limited evidence of its validity (e.g., Darley and Smith 1993). Another justifiable selection would be the Madden, Allen, and Twible (1988) scale since it appears that more work has been conducted to validate it than any other (e.g., Machleit, Allen, and Madden 1993; Madden, Allen, and Twible 1988). For bi-dimensional measurement of A_{ad} , a strong contender is less obvious. However, consideration should be given to the Burton and Lichtenstein (1988) scales because greater evidence of the pair's discriminant validity is available than for any other pair.

Of course, it is quite possible that a semantic differential itself is a problem under some circumstances and alternative measures are needed. For example, even though Likert-type scales have rarely been used to measure A_{ad} in scholarly research (Bruner and Hensel 1992, 1996), they may be more amenable to measuring the construct in phone surveys and other situations where the statements are read to respondents.

Conclusions

It is clear that there is still little consensus regarding the domain of the A_{ad} construct. Not only is there disagreement about which view of attitude structure is most appropriate but even within a conceptualization there is a considerable lack of measurement standardization. Examination of a decade-and-a-half's worth of research published in seven major journals found 75 multi-item measures of A_{ad} involving 53 different semantic differentials. Although some common content among the scales allowed them to be clustered into three groups, the fact remains that almost half (47%) of them had been used as a set just once. At best, this shows the openness of the field towards alternative measures of one of our most important constructs. At worst, it suggests that there has been too much disregard of previous research, at least as it pertains to measurement. Instead of building upon past measures, it has been more typical to reinvent the wheel and create yet more unique measures that are used one time and ignored thereafter.

The most important recommendation that is made here is for researchers to standardize the measurement of A_{ad} as much as they can and justify what they do. Specifically, it is suggested that researchers use a previously developed scale that shows evidence of validity whenever possible, cite the source, and continue the validation process (standardization). If no scale exists or the alternatives are inadequate, one may have to be adapted or constructed. In such a case the reasoning for the new scale should be clearly delineated and initial evidence of its validity should be provided (justification). At the very least, the unbridled proliferation of new measures and "cherry picking" of items should be discouraged.

References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (1985), *Standards for Educational and Psychological Testing*, Washington, D.C.: American Psychological Association.
- Allen, Chris T. and Thomas J. Madden (1989), "Gauging and Explaining Advertising Effects: Emergent Concerns Regarding Construct and Ecological Validity," in *Cognitive and Affective Responses to Advertising*, Patricia Cafferata and Alice Tybout, eds., Lexington, MA: Lexington Books, 327-351.
- Andrews, J. Craig, Syed H. Akhter, Srinivas Durvasula, and Darrel D. Muehling (1992), "The Effects of Advertising Distinctiveness and Message Content Involvement on Cognitive and Affective Responses to Advertising," *Journal of Current Issues and Research in Advertising*, 14 (Spring), 45-58.
- Baker, Michael J. and Gilbert A. Churchill, Jr. (1977), "The Impact of Physically Attractive Models on Advertising Evaluations," *Journal of Marketing Research*, 14 (November), 538-555.
- Batra, Rajeev and Olli T. Ahtola (1990), "Measuring the Hedonic and Utilitarian Sources of Consumer Attitudes," *Marketing Letters*, 2 (April), 159-170.
- Boles, James and Scot Burton (1992), "An Examination of Free Elicitation and Response Scale Measures of Feelings and Judgments Evoked by Television Advertisements," *Journal of the Academy of Marketing Science*, 20 (Summer), 225-233.
- Bonjean, Charles M., Richard J. Hill, and S. Dale McLemore (1967), *Sociological Measurement: An Inventory of Scales and Indices*, San Francisco: Chandler Publishing Company.
- Bowen, John (1990), "Development of a Taxonomy of Services to Gain Strategic Marketing Insights," *Journal of the Academy of Marketing Science*, 18 (Winter), 43-49.
- Brown, Steven P. and Douglas M. Stayman (1992), "Antecedents and Consequences of Attitude toward the Ad: A Meta-Analysis," *Journal of Consumer Research*, 19 (June), 34-51.
- Bruner II, Gordon C. and Paul J. Hensel (1992), *Marketing Scales Handbook—V. I*, Chicago: American Marketing Association.
- and ——— (1996), *Marketing Scales Handbook—V. II*, Chicago: American Marketing Association.
- Buchholz, Laura M. and Robert E. Smith (1991), "The Role of Consumer Involvement in Determining Cognitive Response to Advertising," *Journal of Advertising*, 20 (Spring), 4-17.
- Buros, Oscar Krisen (1975), *Personality Tests and Reviews II*, Highland Park: The Gryphon Press.

- Burton, Scot and Donald R. Lichtenstein (1988), "The Effects Ad Claims and Ad Content on Attitude Toward the Advertisement," *Journal of Advertising*, 17 (Spring), 3-11.
- _____ and George M. Zinkhan (1987), "An Exploratory Investigation of the Dimensions Underlying Attitude Toward the Ad and Their Effect on Purchase Behavior," paper presented at the Southwestern Federation of Administrative Disciplines, Houston, TX.
- Chattopadhyay, Amitava and Kumal Basu (1990), "Humor in Advertising: The Moderating Role of Prior Brand Evaluation," *Journal of Marketing Research*, 27 (November), 466-476.
- _____ and Prakash Nedungadi (1992), "Does Attitude toward the Ad Endure? The Moderating Effects of Attention and Delay," *Journal of Consumer Research*, 19 (June), 26-33.
- Chun, Ki-Taek, J. T. Barnowe, Sidney Cobb, and John R. P. French, Jr. (1974), *Publication and Uses of Psychological Measures in the 1960s*, Ann Arbor, MI: Institute for Social Research, University of Michigan.
- _____, Sidney Cobb, and John R. P. French, Jr. (1975), *Measures for Psychological Assessment*, Ann Arbor, MI: Institute For Social Research, University of Michigan.
- Churchill, Gilbert A. Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64-73.
- Cox, Dena Saliagas and Anthony D. Cox (1988), "What Does Breed? Complexity as a Moderator of Repetition Effects in Advertisement Evaluation," *Journal of Consumer Research*, 15 (June), 111-116.
- _____ and William B. Locander (1987), "Product Novelty: Does it Moderate the Relationship Between Ad Attitudes and Brand Attitudes," *Journal of Advertising*, 16 (Fall), 39-44.
- Cronbach, Lee J. (1971), "Test Validation," in *Educational Measurement*, 2nd ed., R. L. Thorndike, ed., Washington, D.C.: American Council on Education, 443-507.
- Darley, William K. and Robert E. Smith (1993), "Advertising Claim Objectivity: Antecedents and Effects," *Journal of Marketing*, 57 (October), 100-113.
- Dawes, Robyn M. and Tom L. Smith (1985), "Attitude and Opinion Measurement," in *Handbook of Social Psychology*, 3rd ed., Vol. 1, Gardner Lindzey and Elliot Aronson, eds., New York: Random House, 509-566.
- Donthu, Naveen (1992), "Comparative Advertising Intensity," *Journal of Advertising Research*, 32 (November/December), 53-58.
- Droge, Cornelia (1989), "Shaping the Route to Attitude Change: Central Versus Peripheral Processing Through Comparative Versus Noncomparative Advertising," *Journal of Marketing Research*, 26 (May), 193-204.
- Eagly, Alice H. and Shelly Chaiken (1993), *The Psychology of Attitudes*, Fort Worth, TX: Harcourt, Brace, Jovanovich College Publishers.
- Edell, Julie A. and Richard Staelin (1983), "The Information Processing of Pictures in Print Advertisements," *Journal of Consumer Research*, 10 (June), 45-61.
- Finn, Adam and Ujwal Kayandé (1997), "Reliability Assessment and Optimization of Marketing Measurement," *Journal of Marketing Research*, 34 (May), 262-275.
- Furse, David H. and David W. Stewart (1982), "Standards For Advertising Copytesting: A Psychometric Interpretation," *Journal of Advertising*, 11 (Winter), 30-38, 76.
- Gardner, Meryl Paula (1985), "Does Attitude Toward the Ad Affect Brand Attitude Under a Brand Evaluation Set?" *Journal of Marketing Research*, 22 (May), 192-198.
- Goodstein, Ronald C. (1993), "Category-based Applications and Extensions in Advertising: Motivating More Extensive Ad Processing," *Journal of Consumer Research*, 20 (June), 87-99.
- Hair Jr., Joseph F., Rolph E. Anderson, and Ronald L. Tatham (1987), *Multivariate Data Analysis*, New York: Macmillan Publishing Company.
- Hastak, Manoj and Jerry C. Olson (1989), "Assessing the Role of Brand-Related Cognitive Responses as Mediators of Communication Effects on Cognitive Structure," *Journal of Consumer Research*, 15 (March), 444-456.
- Hill, Ronald Paul (1988), "An Exploration of the Relationship Between AIDS-Related Anxiety and the Evaluation of Condom Advertisements," *Journal of Advertising*, 17 (Winter), 35-42.
- _____ (1989), "An Exploration of Voter Responses to Political Advertisements," *Journal of Advertising*, 18 (Winter), 14-22.
- _____ and Michael B. Mazis (1986), "Measuring Emotional Responses to Advertising," *Advances in Consumer Research*, V. 13, 164-169.
- Holbrook, Morris B. and Rajeev Batra (1987), "Assessing the Role of Emotions as Mediators of Consumer Responses to Advertising," *Journal of Consumer Research*, 14 (December), 404-420.
- Holmes, John H. and Kenneth E. Crocker (1992), "Predisposition in the Comparative Effectiveness of Rational, Functional and Discrepant Appeals for Both," *Journal of the Academy of Marketing Science*, 15 (Winter), 11-18.
- Homer, Pamela M (1990), "The Mediation Role of Attitude Toward the Ad: Some Additional Evidence," *Journal of Marketing Research*, 27 (February), 78-86.
- Jacoby, Jacob (1978), "Consumer Research: A State of the Art Review," *Journal of Marketing*, 42 (April), 87-96.
- Janiszewski, Chris (1988), "Preconscious Processing Effects: The Independence of Attitude Formation and Conscious Thought," *Journal of Consumer Research*, 15 (September), 199-209.
- Kamins, Michael A (1990), "An Investigation into the 'Match-Up' Hypothesis in Celebrity Advertising: When Beauty May be Only Skin Deep," *Journal of Advertising*, 19 (Spring), 4-13.
- _____, Lawrence J. Marks, and Deborah Skinner (1991), "Television Commercial Evaluation in the Context of Program Induced Mood: Congruency Versus Consistency Effects," *Journal of Advertising*, 20 (Summer), 1-14.
- Kellaris, James J., Anthony D. Cox, and Dena Cox (1993), "The Effect of Background Music on Ad Processing: A Contingency Explanation," *Journal of Marketing*, 57 (October), 114-124.
- Keller, Kevin Lane (1987), "Memory Factors in Advertising: The Effect of Advertising Retrieval Cues on Brand Evaluations," *Journal of Consumer Research*, 14 (December), 316-333.
- _____ (1991a), "Cue Compatibility and Framing in Advertising," *Journal of Marketing Research* 28 (February), 42-57.
- _____ (1991b), "Memory and Evaluation Effects in Competitive Advertising Environments," *Journal of Consumer Research*, 17 (March), 463-476.
- Kilbourne, William E. (1986), "An Exploratory Study of the Effect of Sex Role Stereotyping on Attitudes Toward Magazine Advertisements," *Journal of the Academy of Marketing Science*, 14 (Winter), 43-46.
- _____, Scott Painton, and Danny Ridley (1985), "The Effect of Sexual Embedding on Responses to Magazine Advertisements," *Journal of Advertising*, 14 (Summer), 48-56.
- King, Maryon F. and Gordon C. Bruner II (1993), "Evaluation of Consumer-Related Scales: A Structured Schemata Approach," *Office of Scale Research Technical Report #9301*.
- Laczniak, Russell N. and Darrel D. Muehling (1993), "The Relationship Between Experimental Manipulations and Tests of Theory in an Advertising Message Involvement Context," *Journal of Advertising*, 22 (Fall), 59-74.

- Lord, Kenneth R., Myung-Soo Lee, and Paul L. Sauer (1994), "Program Context Antecedents of Attitude toward Radio Commercials," *Journal of the Academy of Marketing Science*, 22 (Winter), 3-15.
- Machleit, Karen A. and R. Dale Wilson (1988), "Emotional Feelings and Attitude Toward the Advertisement: The Roles of Brand Familiarity and Repetition," *Journal of Advertising*, 17 (Fall), 27-35.
- _____, Chris T. Allen, and Thomas J. Madden (1993), "The Mature Brand and Brand Interest: An Alternative Consequence of Ad-Evoked Affect," *Journal of Marketing*, 57 (October), 72-82.
- MacInnis, Deborah J. and C. Whan Park (1991), "The Differential Role of Characteristics of Music on High- and Low-Involvement Consumers' Processing of Ads," *Journal of Consumer Research*, 18 (September), 161-173.
- _____, and Douglas M. Stayman (1993), "Focal and Emotional Integration: Constructs, Measures, and Preliminary Evidence," *Journal of Advertising*, 22 (Winter), 51-65.
- MacKenzie, Scott B. and Richard J. Lutz (1989), "An Empirical Examination of the Structural Antecedents of Attitude Toward the Ad in an Advertising Pretesting Context," *Journal of Marketing*, 53 (April), 48-65.
- _____, and Richard A. Spreng (1992), "How Does Motivation Moderate the Impact of Central and Peripheral Processing on Brand Attitudes and Intentions?" *Journal of Consumer Research*, 18 (March), 519-529.
- _____, Richard J. Lutz, George E. Belch (1986), "The Role of Attitude Toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations," *Journal of Marketing Research*, 23 (May), 130-143.
- Macklin, M. Carole, Norman T. Bruvold and Carol Lynn Shea (1985), "Is it Always as Simple as 'Keep It Simple!'" *Journal of Advertising*, 14 (Winter), 28-35.
- Madden, Thomas J., Chris T. Allen and Jacquelyn L. Twible (1988), "Attitude Toward the Ad: An Assessment of Diverse Measurement Indices Under Different Processing Sets," *Journal of Marketing Research*, 25 (August), 242-252.
- Marks, Lawrence J. and Michael A. Kamins (1988), "The Use of Product Sampling and Advertising: Effects of Sequence of Exposure and Degree of Advertising Claim Exaggeration on Consumers' Belief Strength, Belief Confidence, and Attitudes," *Journal of Marketing Research*, 25 (August), 266-281.
- McQuarrie, Edward F. and David G. Mick (1992), "On Resonance: A Critical Pluralistic Inquiry into Advertising Rhetoric," *Journal of Consumer Research*, 19 (September), 180-197.
- Miller, Darryl W. and Lawrence J. Marks (1992), "Mental Imagery and Sound Effects in Radio Commercials," *Journal of Advertising*, 21 (Winter), 83-93.
- Milligan, Glenn W. and Martha C. Cooper (1985), "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika*, 50 (2), 159-179.
- Miniard, Paul W., Sunil Bhatla, and Randall L. Rose (1990), "On the Formation and Relationship of Ad and Brand Attitudes: An Experimental and Causal Analysis," *Journal of Marketing Research*, 27 (August), 290-303.
- Mitchell, Andrew A. (1986), "The Effect of Verbal and Visual Components of Advertisements on Brand Attitudes and Attitude Toward the Advertisement," *Journal of Consumer Research*, 13 (June), 12-24.
- _____, and Jerry C. Olson (1981), "Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude?" *Journal of Marketing Research*, 18 (August), 318-332.
- Muehling, Darrel D. (1987), "Comparative Advertising: The Influence of Attitude-Toward-the-Ad on Brand Evaluation," *Journal of Advertising*, 16 (Winter), 43-49.
- _____, and Russell N. Laczniak (1988), "Advertising's Immediate and Delayed Influence on Brand Attitudes: Considerations Across Message-Involvement Levels," *Journal of Advertising*, 17 (Winter), 23-34.
- _____, and _____ (1992), "An Examination of Factors Mediating and Moderating Advertising's Effect on Brand Attitude Formation," *Journal of Current Issues and Research in Advertising*, 14 (Spring), 23-34.
- _____, _____, and Jeffrey J. Stoltman (1991), "The Moderating Effects of Ad Message Involvement: A Reassessment," *Journal of Advertising*, 20 (Summer), 30-38.
- _____, and Michelle McCann (1993), "Attitude Toward the Ad: A Review," *Journal of Current Issues and Research in Advertising*, 15 (Fall), 25-58.
- _____, Jeffrey J. Stoltman, and Sanjay Mishra (1990), "An Examination of the Cognitive Antecedents of Attitude-Toward-the-Ad," *Current Issues and Research in Advertising*, 12 (Spring & Fall), 95-117.
- Nunnally, Jum C. and Ira H. Bernstein (1994), *Psychometric Theory*, New York: McGraw-Hill.
- Okechuku, Chike and Gongrong Wang (1988), "The Effectiveness of Chinese Print Advertisements in North America," *Journal of Advertising Research*, 28 (October/November), 25-34.
- Olney, Thomas J., Morris B. Holbrook, and Rajeev Batra (1991), "Consumer Responses to Advertising: The Effects of Ad Content, Emotions, and Attitude toward the Ad on Viewing Time," *Journal of Consumer Research*, 17 (March), 440-453.
- Osgood, Charles E., George J. Suci and Percy H. Tannenbaum (1957), *The Measurement of Meaning*, Urbana: University of Illinois Press.
- Percy, Larry and John R. Rossiter (1992), "Advertising Stimulus Effects: A Review," *Journal of Current Issues and Research in Advertising*, 14 (Spring), 75-90.
- Perrien, Jean, Christian Dussart and Francoise Paul (1985), "Advertisers and the Factual Content of Advertising," *Journal of Advertising*, 14 (Spring), 30-35, 53.
- Peter, J. Paul. (1979), "Reliability: A Review of Psychometric Basics and Recent Marketing Practices," *Journal of Marketing Research*, 16 (February), 6-17.
- _____, (1981), "Construct Validity: A Review of Basic Issues and Marketing Practices," *Journal of Marketing Research*, 18 (May), 133-145.
- Peterson, Robert A. (1994), "A Meta-Analysis of Cronbach's Coefficient Alpha," *Journal of Consumer Research*, 21 (September), 381-391.
- _____, William R. Wilson, and Steven P. Brown (1992), "Effects of Advertised Customer Satisfaction Claims on Consumer Attitudes and Purchase Intention," *Journal of Advertising Research*, 32 (March/April), 34-40.
- Petroshius, Susan M. and Kenneth E. Crocker (1989), "An Empirical Analysis of Spokesperson Characteristics on Advertisement and Product Evaluations," *Journal of the Academy of Marketing Science*, 17 (Summer), 217-225.
- Prakash, Ved (1992), "Sex Roles and Advertising Preferences," *Journal of Advertising Research*, 32 (May/June), 43-52.
- Rentz, Joseph O. (1987), "Generalizability Theory: A Comprehensive Method for Assessing and Improving the Dependability of Marketing Measures," *Journal of Marketing Research*, 24 (February), 19-28.
- SAS Institute, Inc. (1988), *SAS/STAT User's Guide*, Release 6.03, Cary, North Carolina.

- Severn, Jessica, George E. Belch, and Michael A. Belch (1990), "The Effects of Sexual and Non-Sexual Advertising Appeals and Information Level on Cognitive Processing and Communication Effectiveness," *Journal of Advertising*, 19 (Spring), 14-22.
- Sharma, Subhash (1996), *Applied Multivariate Statistics*, New York: John Wiley & Sons, Inc.
- Shimp, Terence A. (1981), "Attitude Toward the Ad as a Mediator of Consumer Brand Choice," *Journal of Advertising*, 10 (Summer), 9-15.
- Singh, Surendra N. and Catherine A. Cole (1993), "The Effects of Length, Content, and Repetition on Television Commercial Effectiveness," *Journal of Marketing Research*, 30 (February), 91-104.
- Smith, Robert E. (1993), "Integrating Information From Advertising and Trial: Processes and Effects on Consumer Response to Product Information," *Journal of Marketing Research*, 30 (May), 204-219.
- and William R. Swinyard (1983), "Attitude-Behavior Consistency: The Impact of Product Trial Versus Advertising," *Journal of Marketing Research*, 20 (August), 257-267.
- Steenkamp, Jan-Benedict E. M., and Hans Baumgartner (1992), "The Role of Optimum Stimulation Level in Exploratory Consumer Behavior," *Journal of Consumer Research*, 19 (December), 434-448.
- Stewart, David W. (1992), "Speculations on the Future of Advertising Research," *Journal of Advertising*, 21 (Fall), 1-18.
- Stout, Patricia A. and Benedicta L. Burda (1989), "Zipped Commercials: Are They Effective?" *Journal of Advertising*, 18 (Winter), 23-32.
- Sujan, Mita, James R. Bettman, and Hans Baumgartner (1993), "Influencing Consumer Judgments Using Autobiographical Memories: A Self-Referencing Perspective," *Journal of Marketing Research*, 30 (November), 422-436.
- Tripp, Carolyn, Thomas D. Jensen, Les Carlson (1994), "The Effects of Multiple Product Endorsements by Celebrities on Consumers' Attitudes and Intentions," *Journal of Consumer Research*, 20 (March), 535-547.
- Varadarajan, P. Rajan (1996), "From the Editor: Reflections on Research and Publishing," *Journal of Marketing*, 60 (October), 3-6.
- Wells, William D. (1964), "EQ, Son of EQ and the Reaction Profile," *Journal of Marketing*, 23 (October), 45-52.
- , Clark Leavitt, and Maureen McConville (1971), "A Reaction Profile for TV Commercials," *Journal of Advertising Research*, 11 (December), 11-17.
- Wiles, Judith A. and T. Bettina Cornwell (1991), "A Review of Methods Utilized in Measuring Affect, Feelings, and Emotion in Advertising," *Current Issues and Research in Advertising*, 13 (Spring & Fall), 241-275.
- Wright, Peter J. (1986), "Reactions to an Ad's Contents versus Judgments of the Ad's Impact," in *Advertising and Consumer Psychology*, Jerry Olson and Keith Sentis, eds., New York: Praeger Press, 108-117.
- Yi, Youjae (1990), "Cognitive and Affective Priming Effects of the Context for Print Advertisements," *Journal of Advertising*, 19 (Summer), 40-48.
- (1993), "Contextual Priming Effects in Print Advertisements: The Moderating Role of Prior Knowledge," *Journal of Advertising*, 22 (Spring), 1-10.
- Zinkhan, George M., William B. Locander, and James H. Leigh (1986), "Dimensional Relationships of Aided Recall and Recognition," *Journal of Advertising*, 15 (Spring), 38-46.
- and Christian F. Zinkhan (1985), "Response Profiles and Choice Behavior: An Application to Financial Services," *Journal of Advertising*, 14 (Fall), 39-44, 51, 66.

Appendix 1
BI-Polar Adjectives Used In A_{ad} Scales

-
- | | |
|--------------------------------------|---|
| 1. good/bad | 28. meaningful/meaningless |
| 2. like/dislike | 29. valuable/not valuable |
| 3. irritating/not irritating | 30. important to me/not important to me |
| 4. interesting/uninteresting | 31. beautiful/ugly |
| 5. inoffensive/offensive | 32. positive/negative |
| 6. trustworthy/untrustworthy | 33. satisfactory/not satisfactory |
| 7. persuasive/not persuasive | 34. entertaining/not entertaining |
| 8. informative/uninformative | 35. original/unoriginal |
| 9. believable/unbelievable | 36. dynamic/dull |
| 10. effective/not effective | 37. refreshing/depressing |
| 11. appealing/unappealing | 38. pleasing/irritating |
| 12. impressive/unimpressive | 39. enjoyable/not enjoyable |
| 13. attractive/unattractive | 40. fun to watch/not fun to watch |
| 14. eye-catching/not eye-catching | 41. helpful/not helpful |
| 15. clear/not clear | 42. useful/not useful |
| 16. favorable/unfavorable | 43. fond of/not fond of |
| 17. fair/unfair | 44. well made/poorly made |
| 18. pleasant/unpleasant | 45. insulting/not insulting |
| 19. fresh/stale | 46. sensitive/insensitive |
| 20. nice/awful | 47. soothing/not soothing |
| 21. honest/dishonest | 48. warmhearted/cold hearted |
| 22. convincing/unconvincing | 49. likely/unlikely |
| 23. complete/incomplete | 50. affectionate/not affectionate |
| 24. well-structured/badly structured | 51. weak/strong |
| 25. agreeable/disagreeable | 52. refined/vulgar |
| 26. tasteful/tasteless | 53. familiar/novel |
| 27. artful/artless | |
-

Appendix 3
Evaluation of Scale Origin

STUDY		SOURCE		Conclusion ⁴
Authors (date)	Scale Name ¹	Authors' Statements ²	Overlap in Item Sets ³	
Andrews et al. (1992)	Aad	Combination of measures (two cites)	75% (Holbrook & Batra 1987)	Modified
Boles & Burton (1992)	Aad	Items used by Holbrook & Batra (1987)	100% (Holbrook & Batra 1987)	Borrowed
Buchholz & Smith (1991)	Aad	Items used by MacKenzie & Lutz (1989)	40% (Keller 1987)	Original*
Burton and Lichtenstein (1988)	cognitive	Items drawn from unspecified literature & Wells (1964)	33% (Kilbourne et al. 1985)	Original*
Burton and Lichtenstein (1988)	affective	Items drawn from unspecified literature & Wells (1964)	25% (Zinkhan & Zinkhan 1985)	Original*
Chattopadhyay & Basu (1990)	Aad	Nothing said about source	60% (Mitchell & Olson 1981)	Modified*
Chattopadhyay & Nedungadi (1992)	Aad	Nothing said about source	75% (Hastak & Olson 1989)	Modified*
Cox & Cox (1988)	Ad eval.	Nothing said about source	100% (Cox & Locander (1987)	Borrowed
Cox & Locander (1987)	Ad eval.	Similar to items used by Mitchell & Olson (1981)	40% (Mitchell & Olson (1981)	Modified
Darley & Smith (1993)	Aad	Cite several sources	75% (Mitchell & Olson)	Modified*
Donthu (1992)	Aad	Cites Holmes & Crocker (1987)	90% (Holmes & Crocker 1987)	Modified*
Droge (1989)	Aad	Cites Mitchell & Olson (1981)	60% (Mitchell & Olson (1981)	Modified*
Gardner (1985)	Aad	Follows Mitchell & Olson (1981)	100% (Mitchell & Olson (1981)	Borrowed
Goodstein (1993)	Aad	Nothing said about source	75% (Holbrook & Batra 1987)	Modified
Hastak & Olson (1989)	Ad eval.	Nothing said about source	75% (Mitchell & Olson 1981)	Modified*
Hill (1988)	Global Aad	Suggests modifying Mitchell & Olson (1981) scale	80% (Mitchell & Olson 1981)	Modified
Hill (1988)	emotional	Suggests borrowing Hill & Mazis (1985) scale	40% (Hill & Mazis 1985)	Modified
Hill (1989)	Global Aad	Suggests modifying Mitchell & Olson (1981) scale	100% (Hill 1988)	Modified
Hill (1989)	emotional	Suggests borrowing Hill & Mazis (1985) scale	100% (Hill 1988)	Borrowed
Holbrook & Batra (1987)	Aad	Nothing said about source	100% (Hill 1988)	Borrowed
Homer (1990)	Aad	Nothing said about source	33% (Mitchell & Olson 1981)	Original
Janiszewski (1988)	Ad eval.	Nothing said about source	50% (Muehling 1987)	Unknown*
Kamins (1990)	Aad	Similar to Smith & Swinyard (1983) & Gardner (1985)	50% (Zinkhan & Zinkhan 1985)	Unknown*
Kamins, Marks, & Skinner (1991)	Aad	Similar to Marks & Kamins (1988)	40% (Cox & Locander 1987)	Original*
Kellaris, Cox, & Cox (1993)	Aad	Similar to Marks & Kamins (1988)	75% (Kamins 1990)	Modified*
Keller (1987)	Aad	Scale taken from MacKenzie, Lutz, & Belch (1986)	86% (Madden, Allen, & Twible 1988)	Modified*
Keller (1991a)	Aad	Cites Edell & Staelin (1983)	60% (Mitchell & Olson 1981)	Modified
Keller (1991b)	Aad	Nothing said about source	100% (Keller 1987)	Borrowed
Kilbourne (1986)	Aad	Nothing said about source	100% (Keller 1987)	Borrowed
Kilbourne (1986)	cognitive	Scale previously tested by Baker & Churchill (1977)	100% (Kilbourne, et al. 1985)	Borrowed
Kilbourne, Paunton, & Ridley (1985)	affective	Scale previously tested by Baker & Churchill (1977)	100% (Kilbourne, et al. 1985)	Borrowed
Kilbourne, Paunton, & Ridley (1985)	cognitive	Each item used by Baker & Churchill (1977)	50% (Baker & Churchill 1977)	Borrowed
Kilbourne, Paunton, & Ridley (1985)	affective	Each item used by Baker & Churchill (1977)	75% (Baker & Churchill 1977)	Modified
Laczniak & Muehling (1993)	Aad	Nothing said about source	78% (Muehling, et al. 1988)	Modified*
Lord, Lee, & Sauer (1994)	Aad (claim)	Nothing said about source	75% (Miniard, Bhatla, & Rose 1990)	Modified*
Lord, Lee, & Sauer (1994)	Aad (non-claim)	Miniard, Bhatla, & Rose (1990) cited in lit review	100% (MacKenzie & Lutz 1989)	Borrowed
Machlett, Allen, & Madden (1993)	Aad	Scale developed by Madden, Allen, & Twible (1988)	100% (Madden, Allen, & Twible 1988)	Borrowed
Machlett & Wilson (1988)	Aad	Nothing said about source	33% (Mitchell & Olson)	Original*
MacInnis & Park (1991)	Aad	Nothing said about source	60% (Keller 1987)	Modified*
MacInnis & Stayman (1993)	Aad	Nothing said about source	60% (MacInnis & Park 1991)	Modified*
MacKenzie & Lutz (1989)	Aad	Nothing said about source	50% (Cox & Locander)	Unknown
MacKenzie & Spreng (1992)	Aad	Nothing said about source	100% (MacKenzie & Lutz 1989)	Borrowed
Macklin, Bruvold, & Shea (1985)	Aad	Nothing said about source	11% (Mitchell & Olson)	Original*
Madden, Allen, & Twible (1988)	Ad eval.	Several cites mentioned	50% (Cox & Locander 1987)	Original
McQuarrie & Mick (1992)	Ad liking	Nothing said about source	50% (Cox & Locander 1987)	Modified*
Miller & Marks (1992)	Aad	Cite Mitchell & Olson (1981)	100% (Mitchell & Olson 1981)	Borrowed
Miniard, Bhatla, & Rose (1990)	Aad (claim)	Nothing said about source	50% (Burton & Lichtenstein 1988)	Modified*

Appendix 3 Evaluation of Scale Origin (Continued)

STUDY	Scale Name ¹	Authors' Statements ²	SOURCE	Conclusion ⁴
Miniard, Bhatia, & Rose (1990)	Aad (non-claim)	Nothing said about source	75% (Holbrook & Batra 1987)	Modified*
Miniard, Bhatia, & Rose (1990)	Aad	Nothing said about source	80% (Mitchell & Olson 1981)	Modified*
Mitchell (1986)	Aad	Follows Mitchell & Olson (1981)	100% (Mitchell & Olson 1981)	Borrowed
Mitchell & Olson (1981)	Aad	Nothing said about source	—	Original
Muehling (1987)	Aad	Cites 11 sources for items	25% (Mitchell & Olson 1981)	Original*
Muehling & Laczniak (1988)	Aad	Items taken from Wells, Leavitt, & McConville (1971)	43% (Zinkhan & Zinkhan 1985)	Original
Muehling & Laczniak (1992)	Aad	Patterned after others (two cities)	100% (Muehling & Laczniak 1988)	Borrowed
Muehling, Laczniak, & Stoitman (1991)	Aad	Similar to Madden, Allen, & Twibble (1988)	100% (Muehling & Laczniak 1988)	Borrowed
Muehling, Stoitman, & Mishra (1990)	Aad	Similar items used in several studies ((three cities)	75% (Zinkhan & Zinkhan 1985)	Modified*
Okechuku & Wang (1988)	Aad	Follow Baker & Churchill (1977)	100% (Baker & Churchill 1977)	Borrowed
Okechuku & Wang (1988)	cognitive	Follow Baker & Churchill (1977)	80% (Baker & Churchill 1977)	Modified
Olney, Holbrook, & Batra (1991)	utilitarian	Components taken from Batra & Ahtola (1990)	17% (Kilbourne et al. 1985)	Original*
Olney, Holbrook, & Batra (1991)	hedonism	Components taken from Batra & Ahtola (1990)	17% (MacKenzie & Lutz 1989)	Original*
Perrien, Dussart, & Paul (1985)	cognitive	Items drawn from unspecified literature	40% (Baker & Churchill 1977)	Original*
Perrien, Wilson, & Paul (1985)	utilitarian	Items drawn from unspecified literature	14% (Baker & Churchill 1977)	Original*
Peterson, Wilson, & Brown (1992)	utilitarian	Nothing said about source	29% (Miniard, Bhatia, & Rose 1990)	Original*
Petroshius & Crocker (1989)	cognitive	Suggests consistency w/Baker & Churchill (1977)	100% (Baker & Churchill 1977)	Borrowed
Petroshius & Crocker (1989)	utilitarian	Suggests consistency w/Baker & Churchill (1977)	100% (Okechuku & Wang 1988)	Borrowed
Prakash (1992)	Aad	Cites Gardner (1985) & Mitchell (1986)	100% (Mitchell & Olson 1981)	Borrowed
Singh & Cole (1993)	Aad	Nothing said about source	38% (Muehling 1987)	Original*
Smith (1993)	Aad (eval.)	Scale adapted from Madden, Allen, & Twibble (1988)	46% (Madden, Allen, & Twibble 1988)	Modified*
Steenkamp & Baumgartner (1992)	Aad	Cite MacKenzie & Lutz (1989)	100% (MacKenzie & Lutz 1989)	Borrowed
Sujan, Bettman, & Baumgartner (1993)	Aad	Cite MacKenzie, Lutz, & Belch (1986)	75% (MacKenzie & Lutz 1989)	Modified
Tripp, Jensen, & Carlson (1994)	Aad	Nothing said about source	100% (Steenkamp & Baumgartner 1992)	Borrowed
Yi (1990)	Aad	Cite Mitchell & Olson (1981)	100% (Mitchell & Olson 1981)	Borrowed
Yi (1993)	Aad	Nothing said about source	100% (Mitchell & Olson 1981)	Borrowed
Zinkhan & Zinkhan (1985)	cognitive	Nothing said about source	100% (Mitchell & Olson 1981)	Borrowed
Zinkhan & Zinkhan (1985)	utilitarian	Nothing said about source	0%	Original*
Zinkhan & Zinkhan (1985)	utilitarian	Nothing said about source	33% (Baker & Churchill 1977)	Original*

¹ Scale Name: How the authors themselves referred to the measure.

² Authors' Statements: A paraphrase of what, if anything, the authors said about the source of their scale or its items.

³ Overlap in Item Sets: Percent of items in two sets that are held in common. Correspondence was reported for the scale (published at least a year earlier) with which the degree of overlap was highest. Examination of item sets was made with all of the other scales in the domain of review as well as with the sources cited by the authors. However, final comparisons were limited to measures of the same construct (e.g., Aad) and did not consider the overlap of items used in measuring different constructs (e.g., Ab).

⁴ Conclusion: Complete overlap in item sets between two studies was interpreted to mean that the later study borrowed from the earlier one. For those scales where the overlap was between 50% and 100% the tendency was to assume that the later study modified the earlier scale. An overlap of less than 50% suggested originality given that the majority of the scale's contents had not been used previously in any known study. When exactly half a scale's items was held in common with a previously published measure the nature of its source was based upon other information. If the other information was itself inconclusive then no conclusion regarding the scale's origin could be safely drawn. Exceptions to these decision rules were possible when the authors' statements and/or development work indicated something different had occurred. The asterisk (*) indicates that the set of items was not found to have been used before or since in any known study.