Building an Interval Scale with Verbal Qualifiers for Measuring Attitude toward the Ad (A_{ad})

Dr. Marc Roy
Professor of Marketing
Département of Management
Université du Québec à Rimouski
300 Allée des Ursulines, Rimouski, Québec, Canada G5L 5G2

Abstract

For many years a large number of researchers have studied attitude toward the ad (A_{ad}) and various scales have been used to measure it. It has been shown that the most popular scales, like Likert or Semantic Differential, may bring about validity issues, because those using them may not perceive the distances between the response categories presented to them as being equal. This study describes a scale building procedure that avoids violating the equal distance assumption required for parametric statistical procedures and tests.

Keywords: Attitude toward the ad; measurement; scale; validity

1.Introduction

In a study focussing on commercial breaks placement, using a sample of students enrolled in business administration at a Canadian University, Roy (2013) showed that the way a commercial interrupts a TV program will influence a viewer's felt mood that will, in turn, affect his reactions to an advertisement. One of the variables thought of as an intermediary between commercial break placement and felt mood was attitude toward the ad (A_{ad}). For the last forty years A_{ad} has been receiving a large amount of attention as a mediator of advertising effects(e.g. MacKenzie, Lutzand Belch, 1983; Homer, 1990; Wahid and Ahmed,2011; Yasin, Anwar and Sajid 2013).Since the main focus of the experiment was not to measure A_{ad}, Roy (2013) chose, for practical reasons, to use a one-dimensional scale.

This decision was based on a meta-analysis by Brown and Stayman (1992). They compared the measurement methodologies used in 47 early nineties researches on A_{ad}. Among them, 36 featured multidimensional scales and for the 11 others, one-dimensional measures were taken. Brown and Stayman (1992) expected that multidimensional scales would produce more reliable results. However, having compared the analyses and results from all these researches, they concluded that no appreciable reliability gain could be obtained by measuring A_{ad} with multidimensional scales.

The large amount of different scales used for measuring A_{ad} in many studies (e.g. Roberts, Laughlin and Wedell, 1999; Derbaix and Poncin, 2005; Bartikowski, Chandon and Gierl, 2007) have shown that employing verbal rating items provides several advantages for attitude measurement. Bartikowski, Chandon and Gierl (2007) state that participants to a marketing experiment should be asked to select from a set of verbal qualifiers those that best describe their attitudes towards a stimulus, and then use them for evaluation. In another research, Roberts, Laughlin and Wedell (1999) compared the Likert scale, frequently used for measuring A_{ad}, to the less popular Thurston technique. Their results suggest that researchers should consider using the latter when measuring attitudes, or one of its recently developed item response counterparts.

Surprisingly, in most recent studies focusing on A_{ad} (e.g. Machleitand Sahni, 1992; Raju, Rajagopal and Unnava, 2002; Wahid and Ahmed, 2011; Yasin, Anwar and Sajid, 2013) a majority of researchers have chosen to use Likert or Semantic Differential scales. This is unanticipated since these types of measuring scales can bring about validity issues (e.g. Vermette, 1991). Those validity problems result from the participant's perception of the distances between the choices of answers offered to them. In truth, this type of scale will be considered as perfectly valid only if the participants perceive the distances between the response categories as being the same.

This is generally not the case, since a large amount of studies have shown that visually equal-appearing scales in questionnaires are susceptible to violations of the equal distance assumption most researchers consider required for parametric statistical procedures and tests (e.g. Labovitz, 1967; Traylor, 1983). In addition, Munshi (2014) has shown that the equal interval assumption may not be valid for Likert scales.

Therefore, to eliminate any doubt as to the validity of the scale used to measure A_{ad}, Roy (2013) decided to pick out custom-made verbal rating items for his research. Pras (1976) refers to such items as having verbal qualifiers "at a psychologically equal distance of each other". The method developed by Roy (2013) to select them is a variation on Thurstone's technique, for which a sample of participants coming from the same population used for the main experiment are asked to choose verbal rating items perceived by them as being at an equal distance from each other. Actually, in total, three samples of participants had to be used for the complete elaboration of the measuring scale.

A summary of the entire process employed for the commercial breaks placement study is shown in Figure 1.

Figure 1: Experimental design used for the commercial breaks placement study

STEPS	SAMPLE SIZE (n)	OBJECTIVE
Littérature Review	--	Choosing initial potential verbal rating items
Prel. Experiment 1	31	Rejecting unacceptable items
Prel. Experiment 2	70	Classify and evaluate remaining items
Prel. Experiment 3	34	Evaluation of potential measuring scales
Main Experiment	72	Effects of commercial break on ad efficacy

2. Building a verbal rating scale for measuring A_{ad}

One of the goals of the study on commercial breaks placement was to measure A_{ad}as a mediator of advertising effects. The main experiment took place at l' Université du Québec à Rimouski, using 72 night time adult undergraduate students in business administration. Their age varied from 19 to 48 (M = 23.60; SD = 4.86). The sample consisted predominantly of women (41; 56.94 %).To build the measuring scale a preliminary experiment was done, using a sample of 70 night time adult undergraduate students in business administration but different from those used for the main study. The participants' age varied from 19 to 31 (M = 23.30; SD = 2.60). This sample also consisted predominantly of women (42; 60 %). No significant statistical differences between data coming from the first and second samples were found, leading to conclude that both samples were similar regarding to the age and sex of the participants. The students in the preliminary experiment were given a detailed description of the main research subject, so that when evaluating the potential verbal rating items they would have a clear idea of what was going to be investigated. Because A_{ad} was to be measured with a one-dimensional scale, it was emphasized to the students that the concept being studied had to be thought of as one-dimensional. Obviously, the justification for using samples of students coming from the same program was to make sure that the verbal items picked would be understood in the same way by both groups.

Prior to these two experiments, a list of 35verbal rating items had been produced by reviewing a large number of recent studies focusing on A_{ad}. These35 items were submitted to another group of 31 students coming from the same program as above, this to make sure that they were all worded similarly, that they did not differ in structure and that they meant the same thing to all the students.

For every evaluated item a choice of four short definitions was offered and the participants had to pick the one closest to their own perception of its meaning. The session ended with an informal discussion. This led to 20 verbal qualifiers finally being kept. They are shown, in alphabetical order, in Table 1.

Table 1: Potential verbal rating items to be used for measuring A_{ad}

ACCEPTABLE	AMAZING
BAD	DREADFUL
EXCELLENT	EXTREMELY POOR
GOOD	GOOD ENOUGH
GREAT	NOT SO GOOD
NOT VERY GOOD	OKAY
POOR	PRETTY GOOD
QUITE GOOD	REALLY BAD
REALLY GOOD	VERY GOOD
VERY POOR	WONDERFUL

The procedure employed in the next experiment, used to evaluate these 20 verbal qualifiers, went as followed:

1. Each student was handed a randomly sorted list of the twenty verbal items shown in Table1.
2. The students were asked to classify these items in increasing order of intensity, going from the very negative to the very positive.
3. The students then had to award a value, going from 0 to 10, to each of the previously classified items. These numerical values (0 to 10) were picked because they were to represent the intervals of the scale used for the main experiment.

3.Results

Ranking, mean and standard deviation for each of the 20 verbal items appear in Table 2. It should be noted that only those in bold type have been selected to be included in the measuring scale.

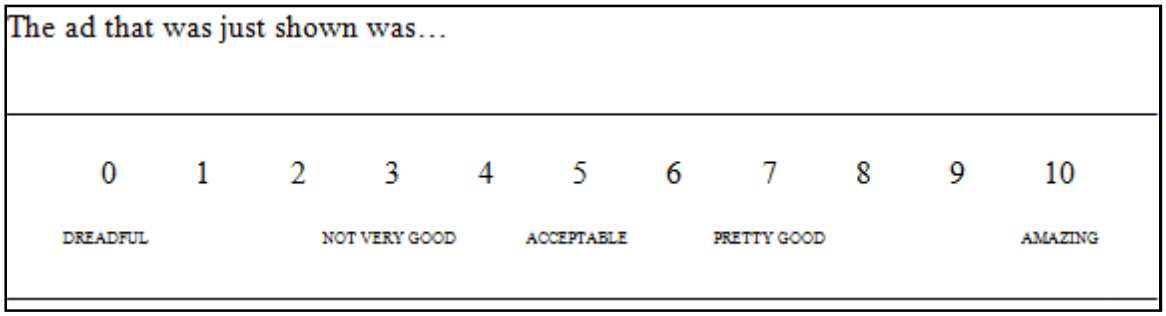
Table 2: Question and A_{ad} Measuring Scale used in main experiment

VERBAL ITEM	MEAN	STANDARD DEVIATION
DREADFUL	**0.08**	**0.25**
REALLY BAD	0.80	0.60
EXTREMELY POOR	1.23	0.67
BAD	1.54	0.92
VERY POOR	1.86	0.79
NOT SO GOOD	2.70	1.06
POOR	2.75	0.96
NOT VERY GOOD	**3.13**	**0.83**
OKAY	4.82	1.09
ACCEPTABLE	**4.86**	**0.86**
GOOD ENOUGH	5.40	1.33
QUITE GOOD	5.61	1.15
GOOD	5.76	0.86
PRETTY GOOD	**7.02**	**0.87**
VERY GOOD	7.22	1.18
REALLY GOOD	7.49	1.01
EXCELLENT	8.53	0.95
WONDERFUL	8.83	1.13
GREAT	9.00	1.16
AMAZING	**9.44**	**1.08**

In order to be picked an item needed to meet two conditions, that is:

a) Its average score had to be very close to the numbers displayed on the scale.
b) Its standard deviation had to be small.

Figure 2: Question and A_{ad} Measuring Scale used in main experiment



Using yet another group of students (n = 34) coming from the same program, a variety of scale versions featuring different numbers of verbal items were evaluated. It was finally decided to adjoin only five verbal qualifiers to the numerical values, this so not to overfill the measuring scale. Figure 2 shows the resulting A_{ad} measuring scale used for the main experiment. To comprehend the motivation behind the specific wording of the question attached to the scale, one should take into account that so to maximize the effects of the treatment studied in the main experiment, it was judged more appropriate that the questionnaire be filled out immediately after the commercial break.

4. Discussion

Even though it has been shown (e.g. Pras, 1976; Roberts, Laughlin and Wedell, 1999; Bartikowski, Chandon and Gierl, 2007; Munshi, 2014) that using a method for building verbal items scales perceived as being at an equal distance from each other provides several advantages for attitude measurement, many researchers still prefer using Likert or Semantic Differential scales. The main reason given is that the Turnstone technique, or a variation of it, is complex and time consuming.

The main objective of this paper was to demonstrate that it is possible to build a scale specifically adapted to the studied variable and the surveyed population, and that this can be done by implementing a relatively simple procedure. The main benefit of this approach is to greatly diminish doubts as to the scale's validity. Although the modus operandi described in this paper may seem somewhat cumbersome, preparing the measuring scale shown in Figure 1 took only a few days. The resulting scale was later used for the main experiment and allowed the observation of a significant mediating role of A_{ad} on advertising effects.

Due to its limited scope, the observations made in this study cannot be generalized in any way. On the other hand, it will hopefully inspire other researchers not to rapidly cast aside the possibility of using a procedure based on the Thurstone technique for measuring A_{ad}.

References

Aylesworth, A. B., MacKenzie, S. B. (1998). Context is the Key: The Effect of Program-Induced Mood on Thoughts about the Ad. Journal of Advertising, 27 (2), 17-31.

Brown, S.P. and Stayman, D.M. (1992). Antecedents and Consequences of Attitude toward the Ad: A Meta-Analysis. Journal of Consumer Research, 19, 34-51.

Bruner I., Gordon C. (1998). "Standardization & Justification: Do Aad Scales Measure Up?," Journal of Current Issues & Research in Advertising, 20 (Spring), 1-18.

Derbaix, C. and I. Poncin (2005). La mesure des réactions affectives en marketing : évaluation des principaux outils. Rechercheet Applications en Marketing, 20, 55-75.

Labovitz, S. (1967). Some observations on measurement and statistics. Social Forces. 46: 151–160.

MacKenzie, S. B., Lutz, R. J. and Belch, G. E. (1986). The Role of Attitude toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations, Journal of Marketing Research, Vol. 23, No. 2 (May, 1986), pp. 130-143.

Machleit, K. A., Sahni, A. (1992).The Impact of Measurement Context on the Relationship Between Attitude Toward the Ad and Brand Attitude For Familiar Brands, Advances in Consumer Research Volume 19, eds. John F. Sherry, Jr. and Brian Sternthal, Provo, UT : Association for Consumer Research, Pages: 279-283.

Munshi, J. (2014). A Method for Constructing Likert Scales, Available at SSRN: http://ssrn.com/abstract=2419366.

Pras, B. (1976).Échelles d'Intervalle à Support Sémantiques, Revue Française du Marketing, n° 61, Mars-Avril 1976, 87-95.

Roy, M. (2013). Effects of Commercial Breaks Placement on Television Advertisement Effectiveness. International Journal of Business and Social Science, 4(9), 73-79.

Traylor,M. (1983). Ordinal and Interval Scaling, Journal of the Market Research Society, 25 (4): 297–303.

Wahid, N. A., Ahmed, M. (2011). The Effect of Attitude toward Advertisement on Yemeni Female Consumers' Attitude toward Brand and Purchase Intention, Global Business and Management Research: An International Journal.

Yasin, S., Anwar, S. and Sajid, M. (2013). Consumer Attitude towards TV Advertising Based Upon ConsumerAge & Gender, Information and Knowledge Management, Vol.3, No.2.