2012

"Pick-any" measures contaminate brand image studies

Sara Dolnicar
University of Wollongong, s.dolnicar@uq.edu.au

John R. Rossiter
University of Wollongong, jrossite@uow.edu.au

Bettina Grun
Johannes Kepler University, bettina@uow.edu.au

Publication Details

Dolnicar, S., Rossiter, J. R. & Grun, B. (2012). "Pick-any" measures contaminate brand image studies. International Journal of Market Research, 54 (6), 821-834.

"Pick-any" measures contaminate brand image studies

Abstract
Brand image measures using the typical "pick-any" answer format have been shown to be unstable (Rungie et al., 2005). In the present study, we find that the poor stability results are mainly caused by the pick-any measure itself because it allows consumers to evade reporting true associations. Using a forced-choice binary measure, we find that stable brand attribute associations are in fact present with much higher incidence (70%), thus outperforming both the measures predominantly used in industry (pick-any, 41%) and academia (7-point scale measure, 59%). Under simulated optimal conditions the forced-choice binary measure leads to 90% stability of brand-attribute associations and is therefore recommended as the optimal answer format for brand image studies.

Keywords
image, pick, any, measures, contaminate, brand, studies

Disciplines
Business | Social and Behavioral Sciences

Publication Details
Dolnicar, S., Rossiter, J. R. & Grun, B. (2012). "Pick-any" measures contaminate brand image studies. International Journal of Market Research, 54 (6), 821-834.

“Pick-any” measures contaminate brand image studies

ABSTRACT

Brand image measures using the typical “pick-any” answer format have been shown to be unstable (Rungie et al., 2005). In the present study, we find that the poor stability results are mainly caused by the pick-any measure itself because it allows consumers to evade reporting true associations. Using a *forced-choice* binary measure, we find that stable brand attribute associations are in fact present with much higher incidence (70%), thus outperforming both the measures predominantly used in industry (pick-any, 41%) and academia (7-point scale measure, 59%). Under simulated optimal conditions the forced-choice binary measure leads to 90% stability of brand-attribute associations and is therefore recommended as the optimal answer format for brand image studies.

KEY WORDS

Brand image measurement, survey research, answer format, answer scale, binary, pick any, forced choice, DLF IIST

Introduction

Consumers' brand-attribute associations recorded with the commercially popular "pick-any" measure are very unstable (Castleberry et al. 1994): the average repeatability of associations on a retest is only 50%. This means: if respondents are asked twice in a row whether they associate Coca Cola with refreshing the first time, the likelihood that they will associate Coca Cola with refreshing again in a second survey is only 50%.

If this low stability were fact, this would mean that brand-image surveys are misleading to the point of being useless because, from the associations recorded on any single survey wave, the brand manager cannot know which associations are valid and which are merely temporary, non-valid reports constructed "on the spot" to satisfy the perceived task requirement imposed by any survey (see especially Krosnick, Narayan and Smith 1996). Low stability, if fact, would also undermine the widely held theory that advertising works by building and reinforcing brand associations (e.g. Keller 2003).

In this study, we demonstrate that the low stability of consumers' brand-attribute associations is mainly due to the use of the commercial market research measure called "pick-any". The pick-any measure (see Figure 1 for an illustration of this answer format) can more technically be described as *free-choice affirmative binary* because respondents are effectively only offered one answer option. They can either tick this one options, indicating that they do associate the brand with that attribute or they can choose not to tick the box next to an attribute, in which case it is assumed that they do not associate the respective brand with the listed attribute.

The pick-any measure is shown to cause respondents to evade many answers that may be true associations and to do so inconsistently, thus producing the low average observed stability. As noted, the average stability is about 50% over a four week interval (actually 49% in the meta-analysis by Rungie et al. 2005). Over an interval of just one week, which is short enough to minimize external influences but long enough to prevent substantial recall of previous answers, the average association's stability with the pick-any measure remains low, at 53% (Dolnicar and Rossiter 2008). We measure stability over the more typical four week interval used in brand tracking studies and record the brands' advertising as a possible intervening variable. We also test two other types of brand-association measures in addition to the free-choice affirmative binary measure. One is the measure widely favored in academic research, the unipolar *7-point scale*. The other one is a *forced-choice* binary measure that was used sometimes by practitioners in the early era of tracking research (see Joyce 1963). Most brand-attribute associations in the questionnaire have been formulated using a specific format of the forced-choice binary measure which is characterized by the fact that the attributes are free of any indication of intensity, so rather than asking respondents whether a laundry detergents "removes tough stains very well" respondents are merely asked whether it "removes tough stains". It assumed that the respondents use their own internal threshold to assess whether or not it removes tough stains well enough to jusitify a "yes" response or not, in which case respondents will answer with "no". This answer format is referred to as *doubly level-free with individual inferred thresholds (*DLF IIST, see Rossiter, Dolnicar and Grün 2010; Rossiter 2011) for details).

For this study, we employ the same stability statistic used in previous studies: the percentage of respondents who say "yes" in both surveys out of those respondents who say "yes" at least once in the two surveys (double positive association rate). "Yes" responses are

defined as: ticks on the free-choice affirmative binary items, where the alternative option is to *not answer* the item; "yes" answers on the forced-choice binary items, where the options are "yes" or "no"; and a score of five or greater on the unipolar ascending 1-to-7 scale items.

The *double positive association* measure emphasizes successful rather than failed associations. It therefore represents a fair criterion for comparison of the pick-any and forced-choice binary measures. It favors the 7-point scale because the consumer can be quite "unstable" by returning scores of "5", "6", or "7" on both interviews and still be counted as "double positive" and thus "stable".

Factors hypothesized to affect brand-attribute association stability

The issue of valid brand image measurement has received some attention in recent years. Two studies have hypothesized possible effects of answer formats on results derived from brand image studies: Dolnicar, Grün and Leisch (2011) compared a forced-choice binary format with a multi-category format concluding that results from the brand image study using the binary format were the same in terms of managerial interpretation and were equally reliable, but saved respondents' time in completing the questionnaire and were perceived by respondents as simpler, thus representing a valid and user-friendly measure for brand image measurement. Dolnicar and Grün (2007) used only the criterion of stability to compare answer formats for brand image measurement and concluded that the forced-choice binary measure produced the most stable answers.

Both previous studies are limited in a critical way, as pointed out by the authors: they use forced-choice binary measures only; they do not study the pick-any measure predominantly used in commercial brand image studies. The key contribution of the present study lies in providing a comparison between the forced-choice binary answer format which has performed well in previous studies with the brand image measure preferred by academics (the unipolar 7-point scale) and the brand image measure preferred in commercial market research (the pick-any measure).

We suspect that the 7-point measure may "overdiscriminate," for low-risk products' ratings especially (Viswanathan, Sudman and Johnson 2004), and thus we hypothesize that the 7-point answer format will produce lower stability than the forced-choice binary answer format. Least stable of all is expected to be the pick-any answer format because of its susceptibility to response evasion (Krosnick, Narayan and Smith 1996). Our hypothesis, and key contribution of this study, accordingly, is as follows:

> ***H1.*** *Double-positive stability will be greatest for the forced-choice binary measure, less for the unipolar 7-point measure, and least for the pick-any measure.*

A number of other measurement factors affect the stability of brand images in two consecutive studies, including consumer involvement with the product category, their familiarity with the particular brand, whether or not it is their preferred brand, the importance of attributes and task ease. All of these factors have been shown to be associated with higher stability of brand images (Dolnicar and Rossiter 2008). We retest these associations to determine whether previous results are supported and to be able to assess the impact of measure effects on brand image stability (specifically the choice of the answer format in the brand image survey) versus effects related to the content of the brand image measurement

task (such as product category involvement, brand familiarity, brand preference). The results of this analysis have direct practical implications: measure effects are fully controlled by the researcher, the other factors are not. Therefore, if the analysis leads to the conclusion that the measure effect is substantial, a simple recommendation can be provided to market researchers, namely to choose the answer format which leads to a more stable brand attribute association by respondents across two repeat measurements.

In the present study, we propose that an additional factor would be influential because we used a four week retest interval: *TV and magazine advertising* for the brand that focuses on a particular attribute or attributes. Such advertising should *reinforce* potentially unstable associations (none of the brands or their campaigns were new, so association *creation* was unlikely) thus increasing stability.

Method

We designed and conducted a brand-image survey in Australia which was administered online to consumers selected from a commercial market research panel. Respondents were asked to assess the six leading brands of laundry detergents along seven attributes found to be most important in a pre-study. Laundry detergents were chosen because they can be regarded as representative of products in brand-image surveys as they "behave" very typically according to the parameters of the stochastic model used previously with the pick-any measure and have been used in previous studies on brand image measurement (see Rungie et al. 2005).

We repeated the survey with the same respondents four weeks later. Respondents were randomly assigned to one of three experimental groups; each group was offered different answer formats in the brand-attribute association task. Respondents were given the same answer format in both survey waves 1 and 2.

Participants

The volunteering rate from the consumer panel for the initial wave of the survey was 29%, which is quite typical for online panels in Australia. Of these respondents, 20% dropped out for the second wave of the survey, which is also quite typical. A comparison of the basic socio-demographic characteristics of non-respondents, respondents, and dropouts indicated that there were significant differences only for the age demographic: consumers under age 35 were significantly less likely to volunteer for the survey and significantly less likely to complete the second wave. The resulting two-wave sample was thus somewhat higher in average age than the nationally representative panel membership.

Given previous findings that respondents who do not have English as their first language significantly reduce brand association stability (Dolnicar and Rossiter 2008) we eliminated the small proportion (4.2%) of respondents who indicated that English was not their first language from the analysis sample. The final sample sizes were 283 for the pick-any measure, 287 for the forced-choice binary measure, and 260 for the unipolar 7-point scale measure.

Measures

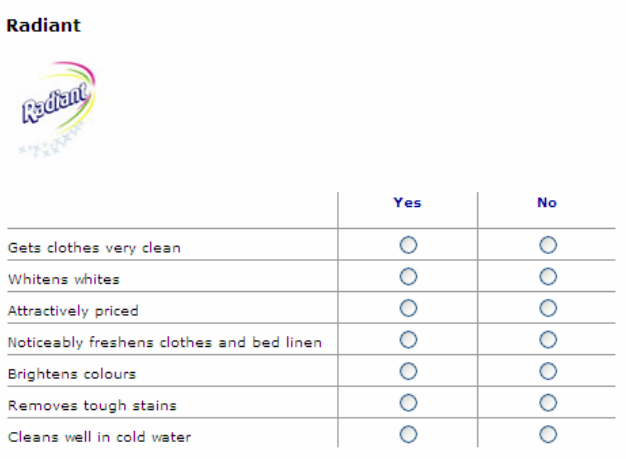
The following brands were used: Omo, Spree, Radiant, Cold Power, Surf and Dynamo. These six brands are the leading brands of laundry detergent in Australia, where the study was conducted. The following attributes were used: Cleans, Freshens, Removes stains, Cold water washing, Whitens, Price, and Brightens. These seven attributes were identified from a small-scale pre-study, with category users, as the most important attributes when selecting a brand of laundry detergent.

Three versions of the questionnaire were designed, one for each measure type. The measures were designed to differ in terms of answer format, but not item content. The exact presentation of the items to respondents is provided in Figure 1.

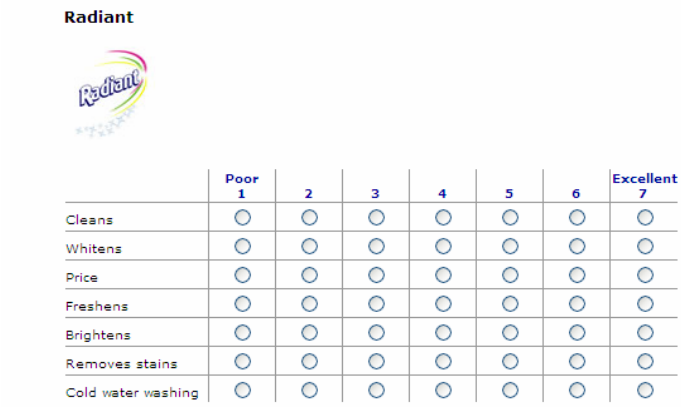
Figure 1: Survey items as presented to respondents (for brand Radiant only)

Pick any

Forced choice binary



Unipolar 7 point



For the analyses, "positive" responses (scored +1) were ticks or "yes" responses on the two binary measures and ratings of "5", "6", or "7" on the 7-point scale; all other responses were scored as zero.

Other variables

In addition to the measure type (answer format) variable, eight other variables were measured as possible predictors of brand-attribute association stability.

Product category involvement was measured by asking respondents to state whether they "were" (scored +1) or "were not" (−1) the "main decider of which brand of laundry powder or liquid is used in your household."

Brand familiarity was measured by asking respondents to state for each brand of detergent whether they have "never heard of it" (scored +1), have "heard of it, but don't know much about it" (+2), are "quite familiar" (+3), or "very familiar" (+4) with the brand.

Brand preference was measured by asking respondents to state their overall opinion of each brand. The answer options were "the single best brand" (scored +2), "one of several very good brands" (+1), "an average brand, not one I would normally buy" (0), "a below average brand" (−1), and "a brand I would not buy under any circumstances" (−2).

Perceived task ease was measured using a 7-point, numbered answer scale with the endpoints labeled "very difficult" (+1) and "very easy" (+7).

Perceived expressability of the rating scale (answer format) was measured using a 7-point, numbered answer scale with the endpoints labeled "not able to express feelings well" (+1) and "able to express feelings well" (+7).

Stated attribute importance (in the laundry detergent product category) was measured using a 3-point answer scale with the answer options labeled "not important at all" (scored 0), "fairly important" (+1), and "very important" (+2).

Brand-attribute advertising for all the brands of laundry detergents was measured by a media monitoring service, which collected all advertisements in TV and magazines during the four week interval. The advertisements were identified by brand and coded by three independent coders in terms of the seven attributes. For each attribute, coders were required to judge the strength to which it was mentioned in the advertisement, coded as "emphasized, that is, promised at an outstanding level" (scored +2), "mentioned, but not at an outstanding level" (+1), or "not mentioned" (0). Rust and Cooil's (1994) PRL (proportional reduction in loss) intercoder agreement statistic for these judgments was .93. Two separate variables were created for the analysis and called simply *TV ads* and *Magazine ads*.

Results

We first report the stability levels of brand-attribute associations obtained from the three alternative measures – pick-any, forced-choice binary, and 7-point. Next we use binary logistic regression with the double positive stability as dependent variable to determine which variables best explain stability and thus point to the causes of instability.

The units of analysis were the brand-attribute associations made on each survey wave. Because multiple observations (six brands by seven attributes = 42) per wave were taken from individual participants in this study we use mixed-effects models for the regression to account for individual differences in stability using a random intercept.

Double-positive stability results

The stability results for the three measures of brand-attribute associations are shown in Table 1. These are reported in terms of three statistics: the mean and two measures of variation, namely, the 95% (two-tailed) confidence interval around the mean, and the observed maximum and minimum values (the range) for the average stability of each brand-attribute combination over respondents. The mean values of double-positive association stability differed significantly from one another at $p < .001$ by pair-wise *t*-tests. The means were exactly in the order hypothesized in H1: greatest for the forced-choice binary measure, less for the unipolar 7-point measure, and least for the pick-any measure.

Table 1: Average response level and observed stability (double-positive associations as a proportion of all positive associations) for the three measure types over the four week retest interval

Measure type	Average response level	**Mean stability**	95% confidence interval	Minimum	Maximum
Forced-choice binary	71%	**.70**	(.67, .73)	.53	.93
Unipolar 7-point	58%	**.59**	(.57, .62)	.38	.79
Pick-any	36%	**.41**	(.37, .44)	.24	.77

Notes: The number of observations is 42 for each answer format measure. The three mean stability estimates differ significantly from one another at $p < .001$.

As expected, the pick-any measure proved to produce very unstable associations with only 41% average double-positive stability. Interpreted in terms of instability, this means that a majority, 59%, of the associations were made in one survey wave but not in the other (four weeks apart). The instability is likely to be due to *evasion* (see Cronbach 1946) encouraged by the pick-any measure which technically allows the respondent to complete the survey

without ticking one single brand attribute association. Evasion was evidenced by the very low *response level* (the aggregate proportion of ticks, averaged over the two survey waves) shown in the first column of Table 1. The average response level on the pick-any measure for the two interview waves was only 36%, which was only half the average positive response level of "yes" answers (71%) that consumers gave when they were using the best measure (the forced-choice binary measure).

The 7-point scale measure fared somewhat better with 59% average stability but was still substantially unstable, with 41% of the positive associations not made consistently. Also, the maximum (most stable) association for the 7-point scale was reported by 79% of consumers, which was no higher than for the pick-any measure, at 77%. Note that the 7-point scale has a significant advantage in the comparison because ticking any of the three positive answer options on the seven point scale was interpreted as a stable response. So instability is higher than the number reported in Table 1 if one would expect respondents to tick the exact same response option (of the seven available) in the two survey waves.

The most stable measure by far, with 70% *average* stability and 93% *maximum* stability, was the forced-choice binary ("yes," "no") measure. This was the only measure to produce associations with *minimum* stability (least stable association) exceeding 50%.

Explanatory results

The main aim of the present study was to try to *explain why* so many brand-attribute associations, even when re-measured over a relatively short interval, are unstable. Hypothesized causes include, firstly, *lack of* the following: involvement with the product category, familiarity with the particular brand, preference for the brand, and importance of the specific attribute. Secondly, as the present study employed a four week retest interval, TV and magazine advertising may be necessary to reinforce associations. Thirdly, the measure type (answer format) used to measure the associations. Lastly, in an attempt to represent the likely psychological process underlying the possible effect of answer format in addition to its objective manipulation, we added two subjective predictors, perceived difficulty of the questionnaire, and how well the answer format enabled participants to express their associations.

For the explanatory analysis, we entered these predictor variables into a binary logistic mixed-effects regression where we included a random intercept for respondents to account for differences in stability between them. The dependent variable was double-positive stability of the association (+1), or not (0), i.e., only one association out of the two measurements. We entered the three measures for the answer format variable as separate predictors using one of them, the pick-any measure, as the base category (included in the intercept) in the regression, so that the regression coefficients for the other two measures represent differences from the base measure. The results are detailed in Table 2. All but three of the explanatory variables were significant predictors at a significance level of $p < .001$; the exceptions were product category involvement ($p = .59$), TV advertising ($p = .44$ and note that magazine advertising had a significant though small *negative* effect on stability, which may mean that ads for *other* brands outweighed those for the test brands), and perceived simplicity of the task ($p = .46$). The only *strong* and significant negative effect on stability was caused by the "pick-any" measure (the intercept in the regression). The notion that advertising will increase stability is not confirmed by the empirical analysis. This finding has

to be interpreted with care, however, because no major advertising campaigns for laundry detergents were launched in this time, so the four week window may not have been enough to reinforce attributes with regular advertising.

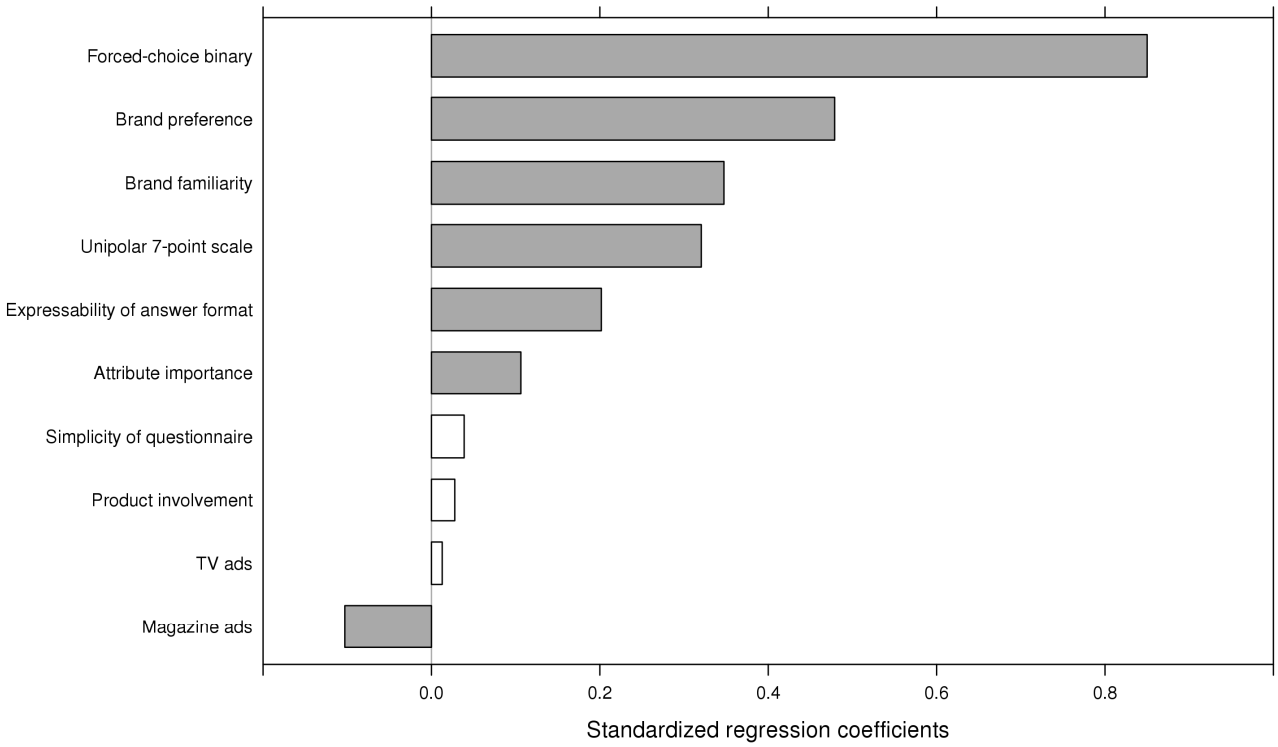
Table 2: Multivariate effects of measure type (intercept: the pick-any measure) and other explanatory variables on the stability of brand-attribute associations via binary logistic regression (dependent variable: double-positive stable or not)

Explanatory variable	Estimated coefficient	S.E.	Wald statistic	Significance (at $p < .001$)
Measure type				
Forced-choice binary	1.722	.123	194.52	sig.
Unipolar 7-point scale	.681	.129	27.95	sig.
Perceived task ease				
Simplicity of Questionnaire	.038	.051	.55	n.s.
Expressability of answer Format	.118	.032	13.40	sig.
Product and brand familiarity				
Product involvement	.034	.062	.29	n.s.
Brand familiarity	.411	.027	224.35	sig.
Brand preference	.667	.029	512.90	sig.
Attribute importance	.207	.036	32.06	sig.
Brand-attribute advertising				
TV ads	.000	.000	.60	n.s.
Magazine ads	– .368	.057	42.50	sig.
Intercept (constant)				
Pick-any measure	– 3.310	.336	97.25	sig.
Nagelkerke R^2	*.34*			

Notes: The number of observations is 23,726.

Figure 2 visualizes the regression results by giving the estimated regression coefficients of the standardized covariates. Regression coefficients which are significantly different from zero are printed in grey, the insignificant ones are printed in white. The lengths of the bars indicate the extent of the influence. If the bar goes the right, the variable increases stability, while the variable decreases stability if the bar goes to the left. As can be seen, using the forced-choice binary answer format in brand image surveys increases stability to the largest extent. As opposed to other factors influencing measurement stability (brand preference, brand familiarity and attribute importance) the choice of survey answer format is entirely in the control of the researcher and thus a measure that can easily be taken to increase the validity of brand image survey data.

Figure 2: Regression coefficients of standardized covariates ordered by their size.



Discussion

Columbia University sociologist Paul Lazarsfeld said that "You never understand a phenomenon until you can make it go away" (Lazarsfeld's pronouncement was reported by Abelson 1995, p. 143, from a 1962 seminar Abelson attended). Short-term instability of consumers' brand-attribute associations should be such a phenomenon. By selecting optimal values of the predictor variables – forced-choice binary measurement and consumers who were familiar with and preferred the brand, rating attributes they regarded as very important, and who perceived the answer format as permitting them to express their feelings well – we observed an *average* stability level of 93%. Thus, we were almost able to make instability "go away" under optimal conditions.

In realistic conditions, stability for the best measure (the forced-choice binary measure) is only about 70%. It is presumed that this is the approximate incidence of valid, pre-existing brand-attribute associations, and the other 30% are assumed to be temporary constructions. The incidence of 70% *remembered* associations may be an overestimate given that some associations, perhaps 10% or so, are easy to construct each time (e.g., an attribute can often be inferred from a literally descriptive or highly suggestive brand name; see Keller, Heckler and Houston 1998) and thus *appear* to be remembered and stable. However, these "ready reconstructions" no doubt have the same positive effect on brand choice as true stable associations. Thus we can estimate as a generalization that, overall, about seven in 10 reported brand-attribute associations are meaningful and three in 10 misleading.

An important practical recommendation follows from this conclusion. The most common consumer sampling method in brand and advertising "continuous tracking" studies is to interview a new sample of consumers on every survey wave (this is the method used by the world's leading tracking suppliers). A new sample each time means that the researcher has no way to distinguish stable associations from unstable ones because there are no repeat interviews and thus no retest of stability. Using a panel or, at least, a semi-panel sampling methodology would solve this problem. A reasonably large sample of individually repeated interviews allows detection of which associations are stable, and therefore presumably valid, and which are false, temporary constructions.

A further practical recommendation is that market researchers should adopt the forced-choice binary measure for brand-image surveys, at least when studying typically low-risk products like laundry detergents. The forced-choice binary measure has a much greater capability to record stable associations than the free-choice practitioner measure and the academically ubiquitous 7-point scale.

References

Abelson, R. P. (1995) *Statistics as principled argument*. Hillsdale, N.J: Lawrence Erlbaum Associates.

Castleberry, S. B., Barnard, N. R., Barwise, T. P., Ehrenberg, A. S. C., & Dall'Olmo Riley, F. (1994) Individual attitude variations over time. *Journal of Marketing Management*, *10*(1-3), pp. 153-162.

Cronbach, L. J. (1946) Response sets and test validity. *Educational and Psychological Measurement*, *6*, pp. 475-494.

Dolnicar, S. & Grün, B. (2007) Question stability in brand image measurement: Comparing alternative answer formats and accounting for heterogeneity in descriptive models. *Australasian Marketing Journal, 15*(2), pp. 26-41.

Dolnicar, S., Grün, B. & Leisch, F. (2011) Quick, simple and reliable: Forced binary survey questions. *International Journal of Market Research*, 53(2), pp. 231-252.

Dolnicar, S., & Rossiter, J. R. (2008) The low stability of brand-attribute associations is partly due to market research methodology. *International Journal of Research in Marketing, 25*(2), pp. 104-108.

Joyce, T. (1963) Techniques of brand image measurement. In *New developments in research – 6th Annual Conference of the Market Research Society* (pp. 45-63). London: Market Research Society.

Keller, K. L. (2003) *Strategic brand management*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Keller, K. L., Heckler, S. E., & Houston, M. J. (1998) The effects of brand name suggestiveness on advertising recall. *Journal of Marketing*, *62*(1), pp. 48-57.

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996) Satisficing in surveys: Initial evidence. *New Directions for Evaluation, 70*, pp. 29-44.

Rossiter, J. R. (2011) *Measurement for the Social Sciences. The C-OAR-SE Method and Why it Must Replace Psychometrics*. New York: Springer.

Rossiter, J. R., Dolnicar, S., & Grün, B. (2010) *The LFFB Comparative Judgment Measure of Brand Attribute Beliefs.* Working Paper, Faculty of Commerce, University of Wollongong, Australia.

Rungie, C., Laurent, G., Dall'Olmo Riley, F., Morrison, D. G., & Roy, T. (2005) Measuring and modeling the (limited) reliability of free choice attitude questions. *International Journal of Research in Marketing, 22*(3), pp. 309-318.

Rust, R. T., & Cooil, B. (1994) Reliability measures for qualitative data: Theory and implications. *Journal of Marketing Research, 31*(1), pp. 1-14.

Viswanathan, M., Sudman, S., & Johnson, M. (2004) Maximum versus meaningful discrimination in scale response: Implications for validity of measurement of consumer perceptions about products. *Journal of Business Research, 47*(1), pp. 108-125.