UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Commerce - Papers (Archive)

Faculty of Business

2010

An initial empirical guide to translating between different answer formats

Sara Dolnicar
University of Wollongong, sarad@uow.edu.au

B. Grun
Vienna University of Economics and Business Administration, bettina@uow.edu.au

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

Publication Details

Dolnicar, S., Grun, B. & Rossiter, J. R. (2010). An initial empirical guide to translating between different answer formats. European Marketing Academy (EMAC) Conference Proceedings. Copenhagen, Denmark: Copenhagen Business School.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

An initial empirical guide to translating between different answer formats

Abstract
Surveys research remains the most popular source of market knowledge. Yet, there is not one established way of how to measure beliefs. Some market research companies offer their respondents five answer options, others seven. Some use middle points on the answer scales, others do not. Some verbalize all answer options, others only the endpoints. The wide variety of answer formats used both by market research companies and academic researchers makes it virtually impossible to compare results across studies. This study offers support by presenting empirical mappings for some of the most commonly used answer formats, thus making comparisons of results easier.

Disciplines
Business | Social and Behavioral Sciences

Publication Details
Dolnicar, S., Grun, B. & Rossiter, J. R. (2010). An initial empirical guide to translating between different answer formats. European Marketing Academy (EMAC) Conference Proceedings. Copenhagen, Denmark: Copenhagen Business School.

This conference paper is available at Research Online: http://ro.uow.edu.au/commpapers/719

An initial empirical guide to translating between different answer formats

Abstract
Surveys research remains the most popular source of market knowledge. Yet, there is not one established way of how to measure beliefs. Some market research companies offer their respondents five answer options, others seven. Some use middle points on the answer scales, others do not. Some verbalize all answer options, others only the endpoints. The wide variety of answer formats used both by market research companies and academic researchers makes it virtually impossible to compare results across studies. This study offers support by presenting empirical mappings for some of the most commonly used answer formats, thus making comparisons of results easier.

Keywords: Survey design, answer formats, translation between answer formats, comparability of study findings

Track 8: Marketing Research and Research Methodology

1 Introduction and Prior Work

Organizations make heavy use of survey research to learn about consumers' beliefs. While repeat surveys by the same organization using the same market research company typically use the same answer format, this is not the case with studies conducted by different organizations, market research companies or academic researchers, making it virtually impossible to compare results across studies. These measurement inconsistencies and the unavailability of any guidance for translating results prevent a body of knowledge from developing, instead leading to many studies producing different sets of numbers.

The aim of this study is to address this problem by providing researchers and users of research studies with a first set of empirical mappings which will enable them to compare results from different answer formats. More specifically, in this study, we will provide mappings which will allow comparison of (1) answer format with different numbers of answer categories; and (2) answer formats with fully verbalized or only endpoint verbalized answer options.

We limit our empirical investigation to the context of brand image measurement. We acknowledge that interactions of construct and answer format occur (Dolnicar and Grün, 2007a; 2009) and results may therefore deviate somewhat for other constructs under study.

Very few studies have been conducted in the past that attempt to map responses from different answer formats onto each other in general, and in the context of brand image measurement in particular. To the authors' knowledge the first study of this kind was conducted by Haley and Case (1979) who evaluated 13 commonly used brand image scales. They compared the answer formats with respect to answer patterns, measured content and concurrent validity and discrimination between brands, concluding that forced choice answer formats as well as answer formats with fully verbalized answer options perform better. Hui and Triandis (1989) compared responses from 5 and 10 point answer formats for Hispanic and Non-Hispanic respondents. However, their research design did not permit mapping across answer formats. The chart provided in the paper, instead shows frequencies of use for each answer option for both formats, indicating that more answer options reduce extreme response style. More recently, mappings between a limited number of answer formats were provided by Dolnicar and Grün (2007a). The study contained measures of two different constructs: behavioral intentions and attitudes. A repeat measurement design on three different answer formats (Full Binary, Metric and Ordinal 7 Point) for each respondent was employed.

2 Data and Methodology

We conducted a large scale experiment using a permission based internet panel. Respondents were asked to complete two brand image questionnaires with approximately a two weeks break between measurements. The two questionnaire versions were identical except for the answer format respondents were offered. This design enabled us to derive individual level mappings, meaning that we actually know how each respondent translates their responses from one answer format to another. Thus, the variation between the two measurements is not due to inter-individual differences in the brand perception or changes in brand perception, because no changes in advertising campaigns or other external events occurred in the marketplace during the two week interval which would have affected brand evaluations.

Please note that it has been demonstrated in the past (Rungie et al., 2005) that brand image measurements are not perfectly stable. We reduced this effect by following the measurement recommendations provided by Dolnicar and Rossiter (2008). Also, this effect can be assumed to affect all experimental conditions in the same way, thus not introducing systematic bias.

Respondents assessed two brands: McDonalds, which is very well known among Australians, and Red Rooster, which is less well known. The five attributes presented to respondents were yummy, fast, cheap, healthy, and convenient. We report mappings between the answer formats explained below. Examples of all answer formats are provided in Fig. 1.

Figure 1: Answer formats

Likert 5 Verbal: *McDonald's is*

	Strongly agree	*Agree*	*Neither agree nor disagree*	*Disagree*	*Strongly disagree*
Yummy	❒	❒	❒	❒	❒
Fast	❒	❒	❒	❒	❒

Likert 5 Endpoints: *McDonald's is*

	Strongly agree				*Strongly disagree*
	+2	*+1*	*0*	*-1*	*-2*
Yummy	❒				❒
Fast	❒				❒

Likert 4 Verbal: *McDonald's is*

	Strongly agree	*Agree*	*Disagree*	*Strongly disagree*
Yummy	❒	❒	❒	❒
Fast	❒	❒	❒	❒

Bipolar 7 Verbal: *McDonald's is*

	very	*rather*	*slightly*	*Neither/nor*	*slightly*	*rather*	*very*	
Yummy	❒	❒	❒	❒	❒	❒	❒	*Yuk*
Fast	❒	❒	❒	❒	❒	❒	❒	*Slow*

Bipolar 7 Endpoints: *McDonald's is*

	Very						*Very*	
	+3	*+2*	*+1*	*0*	*-1*	*-2*	*-3*	
Yummy	❒	❒	❒	❒	❒	❒	❒	*Yuk*
Fast	❒	❒	❒	❒	❒	❒	❒	*Slow*

Bipolar 6 Verbal: *McDonald's is*

	very	*rather*	*slightly*	*slightly*	*rather*	*very*	
Yummy	❒	❒	❒	❒	❒	❒	*Yuk*
Fast	❒	❒	❒	❒	❒	❒	*Slow*

The **Likert 5 Verbal** answer format (Likert, 1932) offers five answer options (incl. a middle point), all of which come with a verbal description. The **Likert 5 Endpoints** answer format contains five answer options (incl. a middle point), but only the endpoints come with a verbal description (*Strongly agree* and *Strongly disagree*). The **Likert 4 Verbal** answer format is the same as the Likert 5 Verbal, except that no middle point is offered. Note that this is not the answer format Likert originally recommended. The **Bipolar 7 Verbal** answer format includes two extremes of the attribute, one at the right and one at the left of the actual answer format. Respondents are then asked to state whether the positive or negative applies and to which extent by ticking one of the seven options, all of which are fully verbally labeled. The middle point is neutral. The **Bipolar 7 Endpoints** format is identical to the bipolar seven point answer format except that only the endpoints are labeled. The **Bipolar 6 Verbal** answer format is the same as the Bipolar 7 Verbal answer format but without a middle point.

Table 1: Experimental design and sample sizes

First measurement	**Second measurement**	**Sample size**
Likert 4 Verbal	Bipolar 6 Verbal	101
Likert 4 Verbal	Likert 4 Verbal	208
Bipolar 6 Verbal	Bipolar 6 Verbal	202
Likert 5 Verbal	Bipolar 7 Verbal	95
Likert 5 Endpoints	Likert 5 Verbal	101
Likert 5 Endpoints	Likert 5 Endpoints	206
Likert 5 Verbal	Likert 5 Verbal	207
Bipolar 7 Endpoints	Bipolar 7 Verbal	103
Bipolar 7 Endpoints	Bipolar 7 Endpoints	203
Bipolar 7 Verbal	Bipolar 7 Verbal	204

The experimental design as well as the sample sizes available for all answer formats included in the study are provided in Table 1. Note that some conditions included two measurements using the exact same questionnaires to enable the calculation of baseline stability (control groups), while others exposed respondents to two different questionnaire to enable mapping of responses. Respondents were randomly assigned to the experimental conditions and were not permitted to participate as respondents in more than one of the experimental groups. The final sample size thus amounts to 1630 respondents.

3 Results

3.1 Mappings on versions with different number of answer options

We compared the Likert 4 Verbal with the Bipolar 6 Verbal and the Likert 5 Verbal with the Bipolar 7 Verbal (Fig. 1). In order to be able to interpret Fig. 1 correctly, we have also computed the base level instability for each of those answer formats. The base level instability gives the percentage of respondents who have not used the same answer options twice in a row when presented with the same answer format. These base instabilities are 29% for Likert 4 Verbal, 35% for Likert 5 Verbal, 52% for Bipolar 6 Verbal and 53% for Bipolar 7 Verbal. These base instabilities in themselves have already practical implications. While most users of such multi-category answer formats argue that they want more than two answer options to capture the finer levels of people's agreement, it appears that the price for these finer levels of responses is low reliability, which fundamentally questions the validity of these measures.

Analyzing the mappings, the following key insights can be gained:
(1) The mapping from 4 to 6 answer options (Fig. 1a) is generally quite consistent with expectations: the extreme options in the four answer format version are split up into the two most extreme options. In the case of the negative response these two contain 74% of all original *Strongly disagree* responses, in the case of the positive responses these two options contain 84% of the original *Strongly agree* option. The same effect occurs for the two middle options of the 4 point format, the only surprising mapping result is that 29% of those who originally stated that they *Disagree* state that the attribute *Slightly* applies, thus effectively switching from a negative to a positive brand attribute association.

(2) The mapping from 5 to 7 answer options (Fig. 1b) leads to similar conclusions: the extreme two options in the 7 point answer format capture 92% of original *Strongly agree* responses and 79% of original *Strongly disagree* options. Switching to the positive side reoccurs: 16% of *Disagree*s move to *Slightly* apply. In addition a substantial amount of movement occurs with respect to the original *Neither Agree Nor Disagree* response.

3.2 Mappings from versions with endpoint labelled to fully labelled answer options

We compared Likert 5 Verbal and Bipolar 7 Verbal where all answer options are verbally labeled with Likert 5 Endpoints and Bipolar 7 Endpoints where only the extremes are verbally labeled. The base level instabilities were 35% for Likert 5 Verbal, 53% for Bipolar 7 Verbal, 46% for Likert 5 Endpoints and 52% for Bipolar 7 Endpoints.

Furthermore, we compared the number of responses that was attracted by the endpoints. The underlying assumption is that if only the endpoints are verbally labelled and if verbal labeling acts as a pointer for respondents, one would expect that more respondents would use endpoints if only those are verbally labelled. This assumption is supported empirically: only 20% use the endpoints for Likert 5 Verbal, as opposed to 27% for the Likert 5 Endpoints version ($\chi^2 = 69$, df = 1, p-value < 0.001) and only 19% use the endpoints on the Bipolar 7

Figure 1: Mapping from Likert 4 Verbal to Bipolar 6 Verbal (Fig. 1a) and from Likert 5 Verbal to Bipolar 7 Verbal (Fig. 1b)

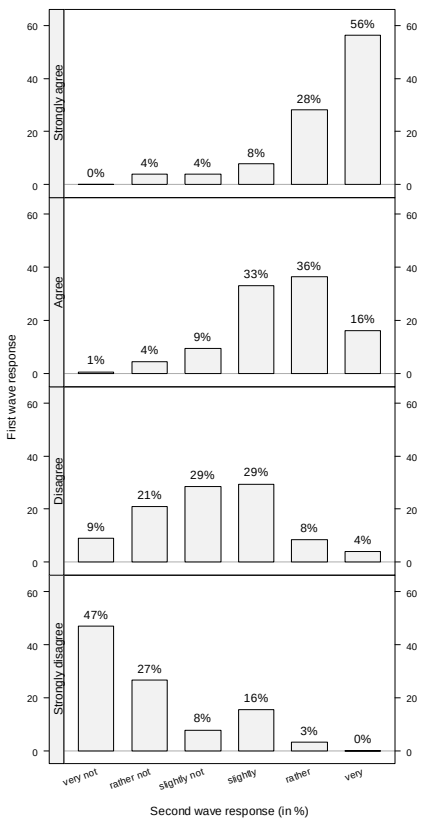

Fig. 1a

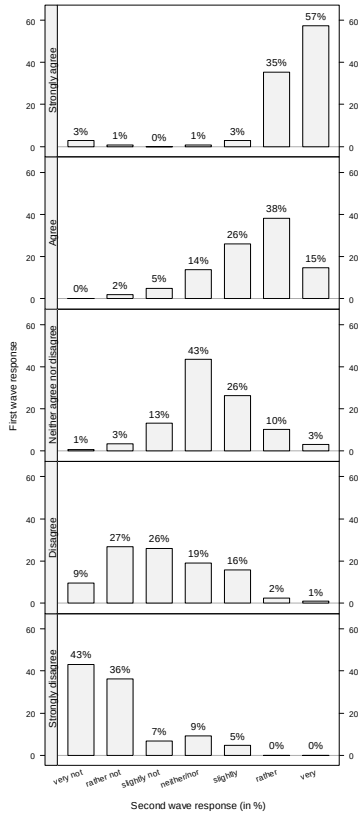

Fig. 1b

Verbal answer format as opposed to 21% in the Bipolar 7 Endpoints version ($\chi^2 = 7.5$, df = 1, p-value = 0.006). The differences are significant for both answer formats, although the difference is smaller for the 7 point answer format.

The mappings are provided in Fig. 2. Overall the switching behavior by respondents when provided with fully verbalized answer formats once and endpoint labeled answer formats amounts to 42% for the 5 point options and 54% for the 7 point options. This means that the switching between the 7 point formats is practically identical to the level for switching that occurs if respondents are presented with the same answer formats twice (the test for proportions for the two base instability levels and the switching rate indicates that they are not statistically significant with p = 0.477; $\chi^2 = 1.5$, df = 2).

The following key insights can be gained from these mappings:
(1) About one third of respondents who were first presented with a Likert 5 Endpoint format and then with a Likert 5 Verbal format moved from *Strongly agree* and *Strongly disagree* to *Agree* and *Disagree*, respectively (Fig. 2a). If, however, they originally stated to *Agree* or *Disagree*, only very few moved to *Strongly agree* (8%) or *Strongly disagree* (13%), respectively. This provides additional empirical support for the previously expressed assumption that endpoint labeled answer formats stimulate extreme responses.

(2) Fig. 2b shows the mapping from Bipolar 7 Endpoints to Bipolar 7 Verbal. The tendency remains the same as described for Likert 5, the only difference is that the level of switching is generally higher, which is in line with the higher base instability rate for this answer format.

Figure 2: Mapping from Likert 5 Endpoints to Likert 5 Verbal (Fig. 2a) and from Bipolar 7 Endpoints to Bipolar 7 Verbal (Fig. 2b)

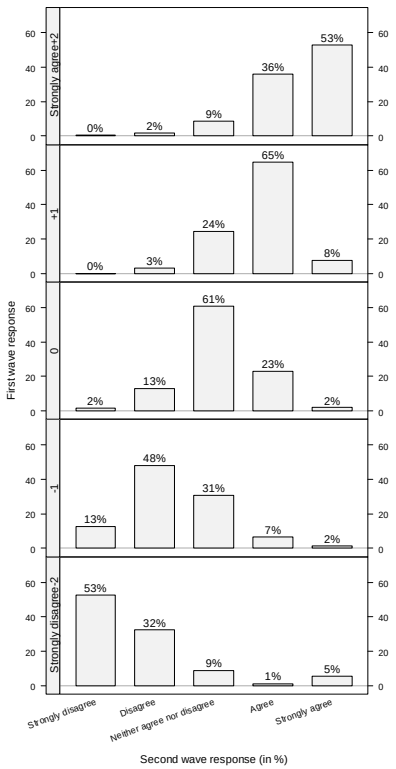

Fig. 2a

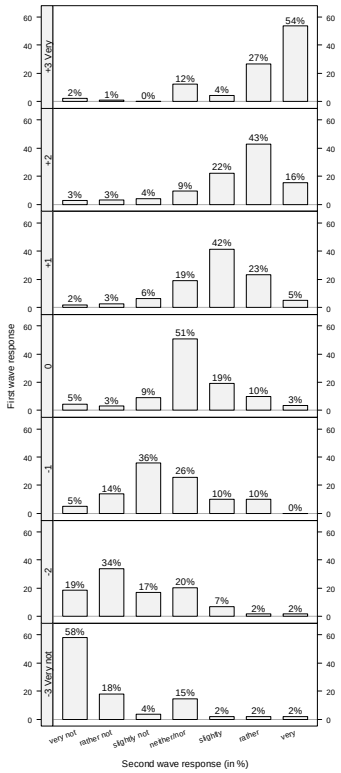

Fig. 2b

4 Conclusions

The aim of this study was to provide empirical mappings of some typical answer formats used to measure beliefs in survey research in order to make it easier for users of empirical study results to compare findings across studies. A number of "behaviors" of answer formats have emerged: (1) When the number of answer options is increased the extreme option on the answer format with fewer options splits into the two most extreme options. (2) A substantial cross-over from the slightest agreement and disagreement levels occurs, both on formats with and without midpoint indicating that respondents choosing those middle points may not hold strong directional beliefs for the brand-attribute associations. And, (3) formats with only the endpoints verbally labeled (as opposed to full verbal labeling) increase the use of extremes.

These findings contribute to our knowledge about the effects of answer format choice in empirical marketing research and provide guidance for the translation of survey findings from studies using different answer formats. The increased understanding of the "behavior" of answer formats also has implications for future choice of answer formats in survey research For example, commonly used 7 point multi-category answer formats (as recommended by Cox, 1980) suffer from a very high base level of instability and may – rather than providing more detail – be capturing more noise, thus making the measurement less valid overall than a simple forced choice binary answer format. This is an important question for future research.

Limitations of this study include that it was conducted in the context of brand image measurement only; replications for other kinds of beliefs are needed. Also, all mappings were based on one particular order of exposure and assume homogeneity among respondents (where, in fact, sub-segments of respondents with different translation functions may exist).

5 References

Cox, E. P. (1980). The Optimal Number of Response Alternatives for a Scale: A Review. Journal of Marketing Research, 17(4), 407-422.

Dolnicar, S. and Grün, B. (2007a). How Constrained a Response: A Comparison of Binary, Ordinal and Metric Answer Formats. Journal of Retailing and Consumer Services, 14(2), 108-122.

Dolnicar, S. and Grün, B. (2007b). Question Stability in Brand Image Measurement - Comparing Alternative Answer Formats and Accounting for Heterogeneity in Descriptive Models. Australasian Marketing Journal, 15(2), 26-41.

Dolnicar, S. and Grün, B. (2009). Does One Size Fit All? The Suitability of Answer Formats for Different Constructs Measured. Australasian Marketing Journal, 17(1), 58-64.

Dolnicar, S. and Rossiter, J.R. (2008). The Low Stability of Brand-Attribute Associations is Partly Due to Measurement Factors. International Journal of Research in Marketing, 25(2), 104-108.

Haley, R.I. and Case, P.B. (1979). Testing Thirteen Attitude Scales for Agreement and Brand Discrimination. Journal of Marketing, 43(4), 20-32.

Likert, R. (1932). A Technique for the Measurement of Attitudes. Archives of Psychology 140, 1-55.

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

6 Acknowledgements

To be added after blind review.