2009

Does one size fit all? The suitability of answer formats for different constructs measured

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

Bettina Grün
Vienna University of Economics and Business Administration, Bettina.Gruen@wu-wien.ac.at

Does one size fit all? The suitability of answer formats for different constructs measured

Abstract

Survey research is used to investigate a variety of different constructs, such as beliefs, behavioural intentions, perceptions, preferences and so on. Despite the wide range of constructs studied by social scientists, the ordinal answer format tends to be used across the majority of survey research studies. We challenge this standard approach in survey research by hypothesizing that the ordinal answer format is not optimal under all circumstances. Instead, we propose that the suitability of answer formats depends on the construct measured. We conduct a repeat measurement study using binary, ordinal and metric answer formats measuring two different constructs: beliefs and behavioural intentions. A clear interaction effect between answer formats and constructs is revealed. This supports the notion that no single answer format is optimal for all research problems, but that some constructs are naturally more suitable for certain answer formats than others. These findings call for increased use of pre-studies to determine the optimal answer format before fieldwork is conducted rather than relying on standard answer formats.

Keywords

marketing measurement, answer formats, binary, ordinal, user-friendliness

Disciplines

Business | Social and Behavioral Sciences

Publication Details

Dolnicar, S. & Grun, B. (2009). Does one size fit all? The suitability of answer formats for different constructs measured. Australasian Marketing Journal, 17 (1), 58-64.

Does one size fit all?

The suitability of answer formats for different constructs measured.

Sara Dolnicar*

School of Management & Marketing,

Marketing Research Innovation Centre (MRIC)

University of Wollongong

Wollongong, NSW 2522, Australia

Telephone: (61 2) 4221 3862, Fax: (61 2) 4221 4154, sara_dolnicar@uow.edu.au

Bettina Grün*

Department of Statistics and Mathematics

Vienna University of Economics and Business Administration

and

Marketing Research Innovation Centre (MRIC)

Augasse 2-6, A-1090 Vienna, Austria

Telephone: (43 1) 31336 5032, Fax: (43 1) 31336 774, Bettina.Gruen@wu-wien.ac.at

*Authors listed in alphabetical order.

Does One Size Fit All?

The Suitability of Answer Formats for Different Constructs Measured.

Abstract

Survey research is used to investigate a variety of different constructs, such as beliefs, behavioural intentions, perceptions, preferences and so on. Despite the wide range of constructs studied by social scientists, the ordinal answer format tends to be used across the majority of survey research studies. We challenge this standard approach in survey research by hypothesizing that the ordinal answer format is not optimal under all circumstances. Instead, we propose that the suitability of answer formats depends on the construct measured.

We conduct a repeat measurement study using binary, ordinal and metric answer formats measuring two different constructs: beliefs and behavioural intentions. A clear interaction effect between answer formats and constructs is revealed. This supports the notion that no single answer format is optimal for all research problems, but that some constructs are naturally more suitable for certain answer formats than others. These findings call for increased use of pre-studies to determine the optimal answer format before fieldwork is conducted rather than relying on standard answer formats.

Keywords: marketing measurement, answer formats, binary, ordinal, user-friendliness

1. Introduction

There would be little resistance among marketing researchers against the statement that different kinds of questions require different answer formats. Yet the ordinal answer format dominates marketing research (Van der Eijk 2001). The vast majority of studies undertaken both by market research companies and by academic researchers use five or seven-point ordinal answer formats in questionnaires, typically Likert-type questions that require respondents to state a certain level of agreement. A simple frequency count of answer formats in one issue of each of the top three journals in the field of marketing (*JMR, JCR* and the *JM*[1]*)* provides empirical support for this statement: 21 articles reported empirical findings based on consumer responses. Of these, 86 percent used ordinal multi-category answer formats.

Research studies comparing answer formats do not support this apparent agreement in the scientific marketing community that ordinal answer formats are the globally optimal choice in questionnaire design. A large number of studies have been conducted to assess the comparative properties of alternative answer formats. Typical criteria used to undertake such comparisons are reliability and validity, structural equivalence, user friendliness and the susceptibility to response styles. Prior work typically used artificial data for such comparative studies or collapsed empirical data with more answer options to fewer options.

Results are controversial. Some studies conclude that if analyses based on means are of interest, binary or trinary answer formats are sufficient and lead to the same results (Lehmann and Hulbert 1972; Loken *et al.* 1987; Preston and Colman 2000; Dolnicar *et al*. 2004), are not less reliable (Bendig 1954; Peabody 1962; Komorita 1963; Komorita and Graham 1965;

---

[1] Journal of Marketing Research, May 2005, Journal of Marketing 69(3) 2005, Journal of Consumer Research 32(1) 2005.

Matell and Jacoby 1971; Jacoby and Matell 1971; Remington *et al*. 1979; Preston and Colman 2000) or valid than multi-category ordinal answer formats (Matell and Jacoby 1971; Jacoby and Matell 1971; Preston and Colman 2000), and do not lead to different findings with regard to the structural equivalence of constructs (Martin *et al.* 1974; Percy 1976).

Contrarily, a number of authors report significant differences with regard to the above criteria, concluding that answer formats offering respondents a larger number of options to choose from lead to better results (studies comparing reliability: Symonds 1924; Nunnally 1967; Jones 1968; Oaster 1989; Finn 1972; Ramsay 1973; studies comparing validity: Loken *et al*. 1987; Hancock and Klockars 1991; studies comparing structural equivalence: Green and Rao 1970). User friendliness and economic efficiency have only been studied by a small number of authors as criteria for comparison between answer formats, again leading to contradictory findings: Jones (1968) concludes that respondents prefer multiple answer options, Dolnicar (2003) and Dolnicar and Grün (2007) find that binary format is more user friendly as it is perceived by respondents to be easier and quicker.

More recently, Dolnicar and Grün (2007) used repeat measurement data to compare alternative answer formats and to investigate how respondents "translate" from one answer format to another. Results indicate that response styles manifest themselves differently on different answer formats. Their analyses of differences in individual mappings between the different answer formats show that the answers on the metric and ordinal answer formats are not comparable and cannot be transformed from one to the other without knowing the response style of the respondents. With respect to managerial interpretations or reliability, no substantial differences were detected in this study.

In addition to the above-mentioned empirical comparison, theoretical discussions and reviews have also been published on answer formats in the past. Not surprisingly, even conceptual and review work does not lead to the same recommendations for survey

researchers. Kampen and Swyngedouw (2000) analyse ordinal scales in detail and postulate that a range of ordinal scales exist which differ in scale properties. They warn empirical researchers of the dangers of inappropriate data assumptions and the ambiguity of interpretations based on frequently ill-defined ordinal formats. Cox (1980), on the other hand, draws the conclusion from his extensive literature review that the seven-point ordinal answer format generally represents a good option, while noting that there is no single optimal answer format for all circumstances. He notes that one of the two main challenges of future work is to establish methods of pre-testing in order to determine which answer format might be most suitable under the given circumstances of the research problem. However, Cox (1980, p. 420) also argues that "scales with two or three alternatives are generally inadequate in that they are incapable of transmitting very much information and they tend to frustrate and stifle respondents."

In sum one can conclude that the search for the optimal answer format is not a new one. An extensive body of work exists in this area of research which is characterised by a range of research questions and approaches. Yet, comparative studies of alternative answer formats have one thing in common: they do not discriminate between differences in the constructs under study. In so doing they implicitly assume the existence of one globally optimal answer format. The main contribution of the present study is to abandon this restrictive implicit assumption and investigate - in general - whether different answer formats are preferable to measure different constructs and – in specific – which of three investigated answer formats appear to best be suited to measure beliefs and behavioural intentions.

In the present study we consequently challenge the assumption that an ordinal answer format represents the generally best option in survey research. We assume that the ability of respondents to correctly differentiate between the grey shades of multiple answer format categories strongly depends on the construct measured. It may be reasonable to ask

respondents to distinguish between several levels of agreement for some constructs in order to be able to measure a value that is as close as possible to their true values of agreement. For instance, respondents may well be able to discriminate reasonably between five agreement levels with the statement "The so-called ecological crisis facing humankind has been greatly exaggerated". This statement developed to measure environmental beliefs is strong and the choice of five agreement levels enables the respondent to state whether they fully or partially agree or disagree with the statement. For other constructs, however, such a fine measurement might not increase the information but the amount of the "noise" in the data which could be, for example, introduced by individual response styles. For instance, if respondents would be asked to indicate how strongly they agree with the statement, "I would bathe my baby in recycled water", it is questionable whether answering with "strongly disagree" or "disagree" would be more of a reflection on the intention to use recycled water for a use with potentially detrimental consequences, or if it would merely reflect a respondent's tendency to strongly agree or mildly agree with statements in general.

Throughout the manuscript we understand the term *binary* or *dichotomous* to indicate a "scale with two mutually exclusive response categories", the term ordinal to indicate a "scale with mutually exclusive and collectively exhaustive categories, as well as the property of order, but not distance or unique origin" and the term metric to indicate "interval and ratio measures" (Cooper and Schindler 2006).

Based on (1) Cox' conclusions, (2) the response style literature which indicates that multi-category ordinal answer formats are susceptible to scale usage heterogeneity, and (3) our assumption that different constructs enable respondents to evaluate responses at different levels of differentiation, we hypothesize that:

H1 The overall use of answer format categories differs for different constructs.

H2 Different people use answer formats differently.

H3 The use of answer format depends on the construct measured.

H4 Ordinal answer formats are perceived as more user-friendly by respondents.

We investigate these hypotheses empirically by comparing responses derived from a binary, a metric and a seven-point ordinal answer format respectively. Measurements using all three answer formats were collected for two different constructs: beliefs and behavioural intentions. In addition, the respondents' evaluations of the user-friendliness of alternative answer formats were recorded.

The findings resulting from the test of the above-listed four hypotheses have major implications for market research. If empirical evidence for the assumption that answer formats are suited differently for different constructs can be provided, two conclusions would have to be drawn, both of which would imply the need for a change of the currently predominant approach of asking questions in survey research. Either (1) pre-studies would have to be conducted before questionnaire development to test which answer format is the most suitable for the construct under study, or, if this is not possible or if all answer formats lead to the same results, (2) it would be preferable to use whichever answer format emerges as the quickest, least complex and cheapest in data collection.

2. Methodology

Repeat measurement data was collected at the University of XXX in three subsequent tutorials held as part of the undergraduate degree offerings at Faculty of Commerce. The order of the answer formats exposure over the three weeks was rotated. Students in one tutorial had the ordinal scale in the first week, the binary in the second and the metric in the third, while the order of answer formats was binary-ordinal-metric and metric-binary-ordinal for the other two tutorials over the three weeks.

Fieldwork was conducted by Research Assistants who were trained before the data collection phase and used standardised verbal instructions when entering the class. Students were told that this survey is part of a research study and that their participation would be much appreciated. Students received no compensation for participating. Given the small class sizes in tutorials, the short duration of the survey (on average 5 minutes completion time) and the personal appeal all students agreed to participate. Missing data resulted from a small number of students not attending all three consecutive tutorials due to sickness.

Student identification numbers were used to match the three questionnaires that contained the same questions using different answer formats: metric, binary (yes-no) and ordinal (seven-point answer format). The questionnaires included questions about two different constructs: behavioural intentions (to use recycled water for different purposes) and beliefs (about environmental protection). The endpoints of the ordinal and metric scales for beliefs were verbalised as “Strongly agree” and “Strongly disagree” whereas the endpoints of the metric answer format were verbalised as “Very likely” and “Very unlikely”. Examples of the scales used are provided in Figure 1. Note that the measurement that was used as the metric scale in the design is a visual analogue scale where respondents are asked to mark on a line the extent to which they agree or disagree with the response. While this is not a truly psychophysical measure, it does specify clearly absolute endpoints as well as distances between any two points along the horizontal line. Consequently our metric measure can be assumed to be clearly distinct from the ordinal measure used in which neither the endpoints are absolute nor the distances between answer options are defined.

Beliefs were measured using a shortened version of the scale known as the New Ecological Paradigm (Dunlap *et al.* 2000) consisting of eight questions. The New Ecological Paradigm Scale in its long (and later shortened) version has been validated and revalidated later by the original authors (Dunlap *et al*. 2000) and has been extensively used in studies of

environmental behaviour to assess different aspects of environmental concern. The following statements were included and will be referred to as the "NEP scale" throughout the article: "The balance of nature is very delicate and easily upset", "When humans interfere with nature it often produces disastrous consequences", "Humans are severely abusing the environment", "The so-called 'ecological crisis' facing humankind has been greatly exaggerated", "If things continue in their present course, we will soon experience a major ecological catastrophe", "Humans have the right to modify the natural environment to suit their needs", "Humans were meant to rule over the rest of nature", and "Plants and animals exist primarily to be used by humans".

Behavioural intentions were measured by asking respondents if they would personally use recycled water for purposes from a list of 13 possible uses of recycled water: Watering the garden, Washing the car, Washing clothes, Cooking, Showering, Taking a bath, Drinking, Toilet flushing, Washing the house, windows or driveways, Watering of garden vegetables and herbs, for use in a Swimming pool, for use in a Fish pond and for Air conditioning.

One sample item for each one of the three answer formats used is provided in Figure 1.

----- Figure 1 about here -----

To assess the user-friendliness and efficiency of alternative answer formats, the starting and finishing times were noted and respondents evaluated each questionnaire with respect to its user-friendliness on a five-point bipolar ordinal answer format. In addition students were asked about their gender and if they mainly speak English with their parents or not. In total, 60 fully completed sets of three questionnaires were available. The repeat-measure nature of the survey is of central importance as it assures that any differences in answer format usage in

dependence of constructs under study is in fact due to the different answer formats and constructs rather than the nature of the sample.

3. Results

All computations and graphics for the empirical analysis have been done using the R statistical software package (R Development Core Team 2007) using package flexclust (Leisch 2006). We decided to use R, an environment for statistical computing and graphics, because it does not only allow for easy routine data analysis (for those familiar with the command line interface) like other standard statistical software packages, but it additionally supports convenient programming and is hence easily extensible and extremely flexible. With the availability of several hundreds of add-on packages written by different members of the R community access to cutting-edge statistical methods is provided.

H1 The overall use of answer format categories differs for different constructs.

The binary and the ordinal seven-point answer format are discrete answer formats. As such, respondents' use of answer categories can be easily compared by determining the absolute and relative frequencies for each category and construct which are provided in Table 1. This comparison indicates that there is no association between construct and the use of the binary answer categories (χ^2=2.41, df=1, p-value=.12) and a significant association between construct and the use of the ordinal answer categories (χ^2=108.32, df=6, p-value<.001): when respondents are asked to assess their behavioural intentions using an ordinal answer format they tend to use the endpoints more frequently. Contrarily, when asked to express their level of agreement with statements relating to environmental beliefs, respondents make more use of

the middle answer categories.

To enable a comparison of the metric (continuous) answer format with the ordinal answer format, the metric answer format is transformed into seven equally spaced intervals. It should be noted, however, that a direct comparison is not possible because respondents' true cut-off points for the translation from the metric to the seven-point answer format are not known. Table 1 provides the absolute and relative frequencies of use for each of the seven categories created from the metric data. Similarly to the ordinal answer format, the endpoints are more frequently used for the behavioural intentions. To compare respondents' patterns of using the metric answer format across the two constructs, a Kolmogorov-Smirnov test for equality of distribution is computed because it avoids imposing cut-off points. Test results indicate a significant difference ($D=.20$, p-value$<.001$) in the way respondents use the metric answer format when asked to evaluate different constructs.

----- Table 1 about here -----

Consequently, H1 cannot be rejected for the metric and the seven-point answer format. In fact, the nature of the differences in using these two answer formats across constructs is similar: respondents use more extreme answer options when asked about behavioural intentions and more middle answer categories when asked about beliefs. For the binary answer format, H1 has to be rejected as no difference in the use of answer categories could be determined across constructs.

H2 Different people use answer formats differently.

There is extensive empirical evidence to suggest that people use answer formats in

different ways. Paulhus (1991, p.17) refers to this as a *response bias*, which is "a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content (i.e. what the items were designed to measure)". In addition he claims that "To the extent that an individual displays the bias consistently across time and situations, the bias is said to be a *response style*". It can consequently be expected that respondents with different response styles would use the three answer formats presented to them in our study in different ways.

Differences in how answer categories are used by individuals are reflected in their answer pattern. An answer pattern for a respondent is the proportion with which he or she uses every single possible answer option (i.e. the relative number of times each answer category is ticked). For the metric answer format – which does not have discrete answer categories - smooth density estimates on a grid with 50-points are determined for each respondent. To avoid confounding the effect of individual answer format use / response styles with the construct effect, the answer patterns of each respondent are determined separately for each construct.

The answer patterns derived are segmented to detect groups of respondents who use answer formats in a similar way. For this purpose, the answer patterns are partitioned using the K-means algorithm using Euclidean distance (Hartigan and Wong 1979) as the underlying measure of dissimilarity. The K-means algorithm is an iterative grouping procedure that aims at minimizing the sum of distances between the answer patterns within each group and maximizing the sum of distances between groups. To ensure detection of a global optimum, the K-means algorithm is repeated with 10 different random initializations and the best solution with respect to the within-sum of distances is reported.

Because natural clusters cannot be expected to exist it is not trivial to choose the optimal number of clusters. A visual inspection of the within-sum of distances for the

different number of groups indicates that a solution with six groups seems to appropriately represent the structure of the binary responses. For the seven-point and the metric answer format four clusters appear to provide the best representation.

The prototypes of the latter two solutions are given in Figure 1. As can be seen two segments are revealed that tend to use the endpoints (either both endpoints or only the positive endpoint). Two other segments clearly avoid the use of endpoints: one of them favours middle categories, whilst the other one prefers answer categories next to the endpoints. These segments are identified for both the metric and the ordinal answer format. To assess whether the same individuals display these response styles, we match the two segmentation solutions. This is possible because we use repeat measures, and answer patterns for the same individuals are included in both the metric and ordinal data set. The matching supports the observation that the correspondence between the two segmentation solutions is high: 60% of the group assignments can be matched (Rand index of .72). A comparison of the group sizes indicates no significant difference (χ^2=1.64, df=3, p-value=.65).

----- Figure 2 about here -----

The cluster memberships were cross-tabulated with the socio-demographic information available for respondents: gender and if the main language spoken with their parents is English. No significant association is detected for any of the tree answer formats (minimum p-value for the six comparisons > 0.18).

Consequently, H2 can not be rejected. Heterogeneity in answer patterns clearly does exist and can be reduced by segmenting respondents into groups with similar answer patterns. This is the case for both ordinal and metric answer formats. In fact, the answer patterns

emerging from both the ordinal and metric data are very similar to each other.

H3 The use of answer format depends on the construct measured.

As opposed to H1, where the use of answer formats was studied across all respondents, H3 takes heterogeneity into account. The groups of respondents with similar answer patterns resulting from the analyses for H2 are used as a starting point for investigating H3.

To assess whether the use of answer formats depends on the construct under study, the assignments of each respondents to an "answer format group" are cross-tabulated with the constructs. This makes it possible to statistically assess whether certain groups occur more or less frequently for one of the constructs using a Pearson's chi-square test.

For the binary answer patterns no significant relationship between answer format usage and construct is detected (χ^2=1.99, df=5, p-value=.85). For the seven-point answer format the association is significant (χ^2=27.96, df=3, p-value<.001), as is the case for the metric answer format (χ^2=48.66, df=3, p-value<.001). Table 2 shows the groups that occur more often for each construct for the seven-point and the metric answer format. As can be seen, answer patterns resulting from the questions relating to behavioural intentions tend to be assigned to groups two and four (those reflecting a higher use of the endpoints).

----- Table 2 about here -----

It can consequently be concluded that individual answer format use differs for the constructs for the ordinal and metric answer format (H3 not rejected) while no difference can be detected for the binary answer format (H3 rejected). In addition the results indicate that the ordinal and the metric answer format are used like a binary answer format for the behavioural

intentions by a significant proportion of respondents.

H4 Ordinal answer formats are perceived as more user-friendly by respondents.

The first measure of user-friendliness (and efficiency) used was the actual time each respondent required to complete the questionnaire. This time was measured in minutes by subtracting the beginning time from finishing time, as provided by respondents. After eliminating a small number of invalid responses (answers with negative durations or durations of more than 20 minutes) 174 observations (97 percent of the total responses) were available for the analysis of user-friendliness.

For the analysis, we included for each measurement whether it was the result of the first, the second or the third measurement (repetition number) because it can be hypothesized that respondents would be quicker in the second and third measurement as they are already familiar with the question. As an indicator for the possible influence of answer format and repetition number, a linear model with the logarithm of duration in minutes or the scores of the perceptions as dependent variable is used. The logarithm is chosen for duration because the distribution of duration is slightly skewed to the right. The influence of repetition and answer format is evaluated using an ANOVA.

The analysis of the time needed to complete the questionnaires point to a difference between repetitions (F-value=23.3, p-value<0.01) and between answer formats (F-value=5.4, p-value<0.01): the binary answer format is completed significantly faster than both the seven-point ordinal (t-value=-3.1, p-value<0.01) and the metric answer format (t-value=-2.6., p-value=0.01). The duration does not differ significantly between the seven-point and the metric answer format (t-value=0.5, p-value=0.62).

In addition to measuring how long it took respondents to complete the questionnaire, respondents were asked to evaluate the scales using a number of specific dimensions relating

to user-friendliness. Respondents were asked to indicate perceived simplicity, perceived pleasantness, perceived speed and perceived ability to express their feelings using a five-point bipolar ordinal answer format. Equidistant scores from 1 to 5 were assigned to the categories where 1 indicates complete agreement and 5 complete disagreement with the question. These scores were used in separate ANOVAs for each subjective evaluation as dependent variables and the repetition number and answer format were used as covariates.

For each subjective evaluation repetition had at least a p-value smaller than 0.1 (Simple: F-value=3.8, p-value=0.02; Pleasant: F-value=6.7, p-value<0.01; Quick: F-value=7.0, p-value<0.01; Express feelings: F-value=2.8, p-value=0.06). The comparative analysis of these items across answer formats indicated that there is no difference with respect to respondents' perceived ability to express feelings (F-value=1.0, p-value=0.36) and the perceived pleasantness (F-value=1.5, p-value=0.23). Differences do, however, exist for perceived quickness (F-value=3.6, p-value=0.03) as the binary answer format is perceived as significantly quicker than the seven-point ordinal answer format (t-value=-2.5, p-value=0.01). No significant differences are observed for simplicity (F-value=2.6, p-value=0.08). However the p-value is rather small and the insignificance might only be due to the small sample size.

Consequently, H4 has to be rejected. This means that respondents do not appear to prefer multi-category ordinal or metric answer formats because they can better express their feelings, as directly suggested by Jones (1968) and indirectly suggested by the fact that multi-category answer formats dominate academic marketing research which is well aware of the negative consequences of user unfriendliness and respondent fatigue on the quality of data collected. This findings has a major practical implication for survey design as it indicated that the standard use of multi-category answer formats is not optimal. Instead, the selection of the answer format must be pre-tested using both criteria of suitability for the construct and user friendliness into consideration to ensure the highest possible data quality.

This result also supports Cox's statement that no single answer format is best under all circumstances, while contradicting his recommendation that two-and three-point answer formats should be avoided. Our study results suggest that the binary answer format is suitable for evaluating behavioural intentions and is not perceived by respondents as more frustrating or stifling. On the contrary, it took less time to complete and was also perceived as quicker by respondents. The ordinal and metric answer formats achieved similar results in terms of the time needed to complete the questionnaire and the perception of user friendliness.

These results suggest that criteria such as user-friendliness, ease and speed of data collection or data requirements for subsequent data analysis methods should play a larger role in the selection process of answer formats, especially if one particular answer format does not appear to be more suitable for a certain construct than others.

4. Conclusions

Since the beginnings of survey research one question of interest to researchers was which answer format should best be used when designing surveys. Consequently an extensive body of knowledge has developed in this area. While studies differ in research questions, approaches and some of the findings, there is a clear tendency of comparative answer format studies to implicitly assume that one singe optimal answer format exists. A good example for such a publication was the review article by Cox (1980), possibly the most prominently published study on the topic in the field of marketing. As Cox states (p. 408) "the purpose of this article is to review the research on the optimal number of response alternatives for a scale." Although he notes in the Conclusions section that there is no single best answer format for all circumstances, he does proceed to conclude that the seven-point ordinal answer format is generally a good option and that answer formats with two or three alternatives are generally not good thus implicitly implying general superiority of certain answer formats.

The aim of the present study was to challenge this implicit assumption of a generally superior scale as well as the common belief among marketing academics that multi-category scales represent one of those generally superior options. This is done by studying the suitability of three alternative answer formats for two constructs typically measured in marketing surveys: beliefs and behavioural intentions.

Results indicate that the same respondents used the same answer formats in a different way when asked to evaluate different constructs. While it appeared that a seven-point ordinal or metric answer format was well suited to capture respondents' beliefs, the patterns of responding to the set of behavioural intentions demonstrated a strong binarisation, indicating that the binary answer format is suitable to capture those responses and can be used without sacrifice in user-friendliness. On the contrary, the binary format led to substantial efficiency gains through reduced completion times. In addition, the comparison of the metric and the seven-point ordinal answer formats indicate that both answer formats trigger a similar answering behaviour from respondents and have a comparable interdependence with constructs.

The choice of the most suitable answer format for a particular research problem is crucial in market research: it affects both the validity of the research (through data quality) and the fieldwork cost. The present study demonstrated the interaction between response formats and constructs measured and illustrates that selecting the most appropriate answer format is not a commonsense problem that can be decided by a researcher alone. Optimally, answer formats should be pre-tested for suitability. This could be achieved by developing a set of suitable answer options for questions to be included in the survey and testing them both qualitatively and quantitatively. A simple qualitative test could include sitting next to respondents as they complete the survey and ask them to share their thoughts about the process of responding with them. Once they completed the survey respondents could be asked for more specific feedback

regarding answer options. For instance, did they find it difficult to choose between the large number of options provided, did they feel restricted by the small number of options provided. They could also be shown a set of answer format alternatives after having completed the test survey and asked directly if any of the other answer formats would have made it easier for them to express their response. A quantitative pre-test could include questionnaire versions with the most suitable answer formats determined in the qualitative phase. A small number of respondents (e.g. 50) could be asked to complete the survey and a frequency analysis of the use of available answer options would provide a reasonable basis to assess how many answer options respondents actually do use.

The limitations of the current study present a number of opportunities for future research.

(1) The sample size in the current study was relatively low.

(2) The sample was limited to students.

(3) The study is limited to two constructs.

(4) The study is limited to three answer formats although even a simple answer format as the binary one can take a number of different forms (for example ticking only “yes” option, offering the respondents the “yes” option before the “no” option or the “no” option before the “yes” option etc.) .

A replication study with a large representative sample of the population would be desirable. Extension studies including larger sets of answer formats and constructs would help shed more light on optimal answer format – construct combinations.

Acknowledgements

To be added after review.

References

Bendig, A.W., 1954. Reliability and the Number of Rating Scale Categories. Journal of Applied Psychology 38, 38-40.

Cooper. D, Schindler. P, 2006, Marketing Research, McGraw-Hill/Irwin.

Cox, E.P., 1980. The optimal number of response alternatives for a scale: A review. Journal of Marketing Research 17, 407-422.

Dolnicar, S., 2003. Simplifying three-way questionnaires - Do the advantages of binary answer categories compensate for the loss of information? ANZMAC CD Proceedings.

Dolnicar, S., Grün, B., 2007. How constrained a response: A comparison of binary, ordinal and metric answer formats. Journal of Retailing and Consumer Services 14, 108-122.

Dolnicar, S., Grün, B., Leisch, F., 2004. Time efficient brand image measurement - Is binary format sufficient to gain the market insight required? CD Proceedings of the 33rd EMAC conference.

Dunlap, R.E., Van Liere, K.D., Mertig A.G., Jones R.E., 2000. Measuring Endorsement of the New Ecological Paradigm: A Revised NEP Scale. Journal of Social Issues 56, 425-442

Finn, R.H., 1972. Effects of some variations in rating scale characteristics on the means and

reliabilities of ratings. Educational and Psychological Measurement 32, 255-265.

Green, P.E., Rao, V.R., 1970. Rating Scales and Information Recovery---How Many Scales and Response Categories to Use. Journal of Marketing 34, 33-39.

Hancock, G.R., Klockars, A.J., 1991. The effect of scale manipulations on validity: Targetting frequency rating scales for anticipated performance levels. Applied Ergonomics 22, 147-154.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-means clustering algorithm. Applied Statistics 28, 100-108.

Jacoby, J., Matell, M.S., 1971. Three-Point Likert Scales Are Good Enough. Journal of Marketing Research 8, 495-500.

Jones, R.R., 1968. Differences in Response Consistency and Subjects' Preferences for Three Personality Inventory Response Formats. Proceedings of the 76th Annual Convention of the American Psychological Association, 247-248.

Kampen, J., Swyngedouw, M., 2000. The Ordinal Controversy Revisited. Quality & Quantity, 34, 87-102.

Komorita, S.S., 1963. Attitude content, intensity, and the neutral point on a Likert scale. Journal of Social Psychology 61, 327-334.

Komorita, S.S., Graham, W.K., 1965. Number of scale points and the reliability of scales. Educational and Psychological Measurement 25, 987-995.

Lehmann, D.R., Hulbert, J., 1972. Are Three Point Scales Always Good Enough? Journal of Marketing Research 9, 444-446.

Leisch, R., 2006. A toolbox for k-centroids cluster analysis. Computational Statistics and

Data Analysis 51, 526-544

Loken, B., Pirie, P., Virnig, K.A., Hinkle, R.L., Salmon, C.T., 1987. The Use of 0-10 Scales in Telephone Surveys. Journal of the Market Research Society 29, 353-362.

Martin, W.S., Fruchter, B., Mathis, W.J., 1974. An investigation of the effect of the number of scale intervals on principal components factor analysis. Educational and Psychological Measurement 34, 537-545.

Matell, M.S., Jacoby, J., 1971. Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. Educational and Psychological Measurement 31, 657-674.

Nunnally, J.C., 1967. Psychometric Theory, McGraw-Hill, New York, 1st edition

Oaster, T.R.F., 1989. Number of alternatives per choice point and stability of Likert-type scales. Perceptual and Motor Skills 68, 549-550.

Paulhus, D.L., 1991. Measurement and control of response bias, in Robinson, J.P., Shaver, P.R. and Wrightsman, L. S. (eds.), Measures of Personality and Social Psychological Attitudes, Academic Press, San Diego pp.17-59.

Peabody, D., 1962). Three components in bipolar scales: direction and extremeness. Psychological Review 69, 65-73.

Percy, L., 1976). An Argument in Support of Ordinary Factor Analysis of Dichotomous Variables, in: Anderson, B. (Ed.), Advances in Consumer Research, Association for Consumer Research, pp. 143-148.

Preston, C.C., Colman, A.M., 2000. Optimal number of response categories in rating scales:

reliability, validity, discriminating power, and respondent preferences. Acta Psychologica 104, 1-15.

R Development Core Team, 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ramsay, J.O., 1973. The Effect of Number of Categories in Rating Scales on Precision of Estimation of Scale Values, Psychometrika 37, 513-532.

Remington, M., Tyrer, P.J., Newson-Smith, J., Cicchetti, D.V., 1979. Comparative reliability of categorical and analogue rating scales in the assessment of psychiatric symptomatology. Psychological Medicine 9, 765-770.

Symonds, P.M., 1924. On the Loss of Reliability in Ratings Due to Coarseness of the Scale. Journal of Experimental Psychology 7, 456-461.

Van der Eijk, C., 2001. Measuring agreement in ordered rating scales. Quality & Quantity 35, 325-341.

1 Tables and Figures

Table 1: Use of scale categories for the three answer formats and two constructs

		Seven-Point Scale							Binary	
		1	2	3	4	5	6	7	Yes	No
Absolute	NEP	54	42	66	88	87	88	53	261	206
	Intentions	169	102	86	68	65	94	195	394	376
Relative	NEP	.11	.09	.14	.18	.18	.18	.11	.56	.44
	Intentions	.22	.13	.11	.09	.08	.12	.25	.51	.49
		Metric								
	from	.00	.15	.30	.44	.58	.72	.87		
	to	.14	.29	.43	.57	.71	.86	1.00		
Absolute	NEP	69	57	54	75	57	95	68		
	Intentions	196	78	64	67	42	67	263		
Relative	NEP	.15	.12	.11	.16	.12	.20	.14		
	Intentions	.25	.10	.08	.09	.05	.09	.34		

Table 2: Cluster assignments given constructs for the metric and the 7-point scale

		Seven-Point Scale				Metric			
		1	2	3	4	1	2	3	4
Absolute	NEP	24	11	17	8	23	11	24	2
	Intentions	3	23	12	22	7	29	3	21
Relative	NEP	.40	.18	.28	.13	.38	.18	.40	.03
	Intentions	.05	.38	.20	.37	.12	.48	.05	.35

Figure 1: Sample answer formats

Binary answer format

	I disagree	**I agree**
The balance of nature is very delicate and easily upset.	❑ [0]	❑ [1]

Ordinal answer format

	1 Strongly disagree	**2**	**3**	**4**	**5**	**6**	**7** Strongly agree
The balance of nature is very delicate and easily upset.	❑[1]	❑[2]	❑[3]	❑[4]	❑[5]	❑[6]	❑[7]

Metric answer format

	Strongly disagree Strongly agree
The balance of nature is very delicate and easily upset.	----------

Figure 2: Answer patterns of the K-means solutions for the metric and the 7-point scale