



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Commerce - Papers (Archive)

Faculty of Business

2012

One legacy of Mazanec: binary questions are a simple, stable and valid measure of evaluative beliefs

Sara Dolnicar

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

Friedrich Leisch

University of Wollongong

Publication Details

Dolnicar, S. & Leisch, F. (2012). One legacy of Mazanec: binary questions are a simple, stable and valid measure of evaluative beliefs. *International Journal of Culture, Tourism and Hospitality Research*, 6 (4), 316-325.

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

One legacy of Mazanec: binary questions are a simple, stable and valid measure of evaluative beliefs

Abstract

Purpose – Academic researchers love multi-category answer formats, especially five- and seven-point formats. More than a decade ago Josef Mazanec concluded that these formats may not be the best choice, and that simple binary-answer options are preferable in some empirical survey contexts. The purpose of the present study is to investigate empirically Mazanec's hypothesis in the context of the measurement of evaluative beliefs relating to fast-food restaurants.

Design/methodology/approach – The authors conducted an online experiment that asked respondents to assess evaluative beliefs relating to fast-food brands using either a forced binary (n = 100) or a seven-point answer format (n = 100). The authors also measured preferences for each of the fast-food restaurants, user friendliness, and recorded the actual completion times for the survey.

Findings – The results indicate that the full binary answer format outperforms the popular seven-point multi-category format with respect to stability, concurrent validity, and speed of completion.

Practical implications – Given the demonstrated strengths of full binary measures, they should be used more by both practitioners and academics when measuring evaluative beliefs.

Originality/value – This study provides empirical evidence of the strong performance of the forced binary-answer format for the measurement of evaluative beliefs, and thus challenges current measurement practice among academics and practitioners.

Keywords

questions, simple, stable, mazanec, valid, legacy, measure, evaluative, beliefs, one, binary

Disciplines

Business | Social and Behavioral Sciences

Publication Details

Dolnicar, S. & Leisch, F. (2012). One legacy of Mazanec: binary questions are a simple, stable and valid measure of evaluative beliefs. *International Journal of Culture, Tourism and Hospitality Research*, 6 (4), 316-325.

One legacy of Mazanec: binary questions are a simple, stable and valid measure of evaluative beliefs

Sara Dolnicar¹, Friedrich Leisch²

This research was funded by the Australian Research Council through the ARC Discovery (DP0878423) and Linkage International (LX0881890 and LX0559628) grant schemes. Authors listed in alphabetical order.

¹ Sara Dolnicar, Institute for Innovation in Business and Social Research, University of Wollongong, Wollongong, NSW 2522, Australia. Tel.: +61 2 4221 3862; fax: +61 2 4221 4210.

² Friedrich Leisch, Institut für Statistik, Ludwig-Maximilians-Universität, Ludwigstraße 33, D-80539 München. Tel.: +49 89 2180 3165, fax: +49 89 2180 5308.

E-mail addresses: sarad@uow.edu.au (S. Dolnicar), Friedrich.Leisch@lmu.de (F. Leisch).

Abstract

Purpose — Academic researchers love multi-category answer formats, especially five- and seven-point formats. More than a decade ago Josef Mazanec concluded that these formats may not be the best choice, and that simple binary-answer options are preferable in some empirical survey contexts. The present study empirically investigates Mazanec's hypothesis in the context of the measurement of evaluative beliefs relating to fast-food restaurants.

Design/methodology/approach – We conducted an online experiment that asked respondents to assess evaluative beliefs relating to fast-food brands using either a forced binary (n = 100) or a seven-point answer format (n = 100). We also measured preferences for each of the fast-food restaurants, user friendliness, and recorded the actual completion times for the survey.

Findings — Results indicate that the full binary-answer format outperforms the popular seven-point multi-category format with respect to stability, concurrent validity, and speed of completion.

Practical implications – Given the demonstrated strengths of full binary measures, they should be used more by both practitioners and academics when measuring evaluative beliefs.

Originality/value – This study provides empirical evidence of the strong performance of the forced binary-answer format for the measurement of evaluative beliefs, and thus challenges current measurement practice among academics and practitioners.

Keywords: Survey design, Answer format, Binary, Multi-category, Stability, Concurrent validity, Speed of completion, Measuring evaluative beliefs

Introduction

Survey research represents one of the key bases for knowledge development and market intelligence in academic and applied tourism research. To ensure that valid conclusions are derived from survey data, designing questionnaires in a way that minimizes measurement error is critical. A range of factors affects the quality survey data, including sampling strategy, length of the questionnaire, wording of the questions, and answer options offered to respondents.

This study focuses on answer options, and challenges the assumption that multi-category answer formats are always preferable to binary formats. Although practically no work in tourism research exists that specifically discusses the advantages and disadvantages of different answer formats, our review of empirical tourism studies reveals an implicit belief held by most tourism researchers: that multi-category formats are superior. Among all empirical studies published in the past five years in the *Journal of Travel Research*, 83 percent use multi-category formats to measure people's beliefs. The two most popular answer formats are the five- and seven- point formats, with approximately one-third of studies using each, followed by the binary and pick-any format, which together account for less than 20 percent of the measures used. Table 1 provides a full analysis of the review.

Another interesting observation that emerges from the review, and which is not reflected in Table 1, is that researchers rarely provide a justification for the answer format they use. More specifically, in 70 percent of the reviewed studies the authors did not even attempt to provide an explanation or justification for their choice of answer format. In 30 percent of cases they do explain, but the vast majority of those justifications argue that someone else used this same answer format in a previous study. Such justifications reflect copying behavior, rather than considered reasoning why the researchers believe that the chosen answer format is likely to lead to a valid measurement.

Such copying behaviour is often rewarded in the publication process. Typically, studies use multi-category answer formats and reviewers, and do not evoke any concerns. This lack of scrutiny is unfortunate because it removes researchers' incentives to choose a valid answer format based on an assessment of the alternatives and justify their choice in the report.

Table 1
Review of answer format use in empirical tourism studies

Number of answer options	Frequency of use	Percentage
1 (pick any, only YES answer option offered)	10	8
2 (both YES and NO answer options offered)	12	10
4	4	3
5	41	33
6	3	2
7	40	32
8	1	1
9	3	2
10	8	6
11	1	1
100	3	2

Although he never published specifically on this topic, the works of Josef Mazanec do not reflect an unquestioned acceptance of multi-category formats. Instead, he preferred to use simple binary formats for eliciting certain kinds of information from survey respondents, for example, in the area of segmentation studies in tourism. Four authors pioneered *a posteriori* (Mazanec, 2000), or data-driven (Dolnicar, 2004), market segmentation in tourism research. Of those, only Mazanec (1984) clusters respondents on the basis of binary data. Calantone, Schewe, and Allen (1980) use six-point answer formats to elicit importance attributes; Goodrich (1980) uses a seven-point format to elicit benefit attributes; and Crask (1981) uses a five-point format to measure vacation attributes, which are used for the segmentation task.

Mazanec also includes a detailed explanation for his unconventional choice of answer format:

In travel research applications, moreover, we have to cope with a complex product (destinations, package tours) offering the consumer a wide range of benefits from which to choose. With a voluminous battery of rating scales, the measurement process is likely to become onerous and boring to respondents. Since we do not want to endanger the reliability of information collected, we have to simplify the measurement approach. In the author's experience, it is preferable to economize on scale levels rather than on number of benefit items. Measurement of benefits is easiest for the respondent if he is asked only to evaluate a benefit item as being important or not important (Mazanec, 1984, p. 18).

This study empirically tests the hypothesis that the binary-answer format outperforms the most commonly used multi-category answer format — the seven-point format — as a measure of evaluative beliefs in survey research. This study only investigates evaluative beliefs, and consequently, conclusions drawn about comparative performance of answer formats are limited to this field. For other constructs, such as overall attitude, some have argued conceptually (Rossiter, 2011) that binary-answer formats are not appropriate.

We may measure performance in the “answer format competition” in three ways: 1) stability of responses over repeated measurements, 2) concurrent validity of responses with respect to preferences, and 3) user-friendliness, including time required to complete the questionnaire.

This study focuses only on answer formats — not the wording of the questions (which is another common source of validity problems), nor the sample size (which determines precision of results). Specifically, the answer format should be: 1) free of bias, which is a condition for 2) content validity, which in turn is a condition for 3) test-retest reliability (stability), which is a condition of 4) predictive or concurrent validity. This study empirically tests stability and concurrent validity, and also compares user-friendliness, a criterion we view as supplementary. Therefore, if both answer formats perform equally well on stability and concurrent validity, the answer format that is more user-friendly would be preferable. However, we do not endorse compromise on stability or concurrent validity in order to increase user-friendliness. Our hypotheses follow:

- H1** When measuring evaluative beliefs, the binary-answer format outperforms the seven-point format regarding stability over time because it focuses on the direction of the answer and does not confound the direction response with an

intensity response (Peabody, 1962; Komorita, 1963; Albaum, Rogers, Roster & Yu, 2006).

- H2** When measuring evaluative beliefs, the binary-answer format achieves greater concurrent validity than the seven-point format (Bendig, 1954; Dolnicar, 2003; Komorita & Graham, 1965; Matell & Jacoby, 1971a; 1971b; Martin, Fruchter & Mathis, 1974; Peabody, 1962; Schutz & Rucker, 1975).
- H3** When measuring evaluative beliefs, the binary-answer format outperforms the seven-point answer format regarding user-friendliness because it requires less cognitive effort and takes less time to answer (Jones, 1968; Dolnicar, 2003; Dolnicar & Grün, 2007a).

The overall aim of the study is to raise awareness among tourism researchers about the importance of considering carefully which answer options to offer respondents, given that the “major advantage of measurement is taking the guesswork out of scientific observation” (Nunnally & Bernstein, 1994, p. 6). To achieve greater accuracy, the choice of answer format cannot be based on guesswork or habit — it must be justified — and will likely require preliminary qualitative research in order to ensure content validity (Rossiter, 2011).

Prior work

Questionnaire design generally — and specifically, the effects of answer formats — have attracted a substantial amount of attention among researchers over the past decades. The number of different recommendations regarding the optimal number of answer options to use is almost as high as the number of studies that investigate the matter. This is partly because studies use different criteria to assess the performance of alternative answer formats and include answer options with not only different numbers of answer options, but also different labeling and presentation techniques. Consequently, determining any clear consensus between conclusions drawn from prior studies is impossible. We therefore summarize prior work by presenting the key arguments made for and against both multi-category and binary-answer formats.

Arguments in support of multi-category answer formats

The majority of studies that argue in favour of using multi-category answer formats use a measure of internal consistency, often referred to more generally as reliability, and typically use Cronbach’s alpha as the measure. Based on these studies, authors recommend five-point answer formats (Remmers & Ewart, 1941, Lissitz & Green, 1975; Jenkins & Taber, 1977), seven-point answer formats (Symonds, 1924; Oaster, 1989; Finn, 1972; Cicchetti, Showalter & Tyrer, 1985), and 18–24 point formats (Champney & Marshall, 1939).

Chang (1994) challenges this body of research by demonstrating that higher numbers of answer options cause larger response sets, which in turn lead to inflated correlations. The increased levels of internal consistency are thus, at least partially, a statistical artefact. Chang calls for a “separation of method variance from internal consistency” (p. 212). Despite Chang’s arguments against using coefficient alpha, more recent studies advise using this measure as a criterion for comparing answer formats: Preston and Colman (2000) use several criteria, including Cronbach’s alpha,

and conclude that questionnaires should use rating formats with seven, nine or ten answer options.

A small number of studies use a test-retest design and other criteria to compare answer format performance: comparing answer formats on the basis of stability, Boote (1981) recommends using five-point answer formats; using information transmission as the criterion for comparison, Garner (1960) recommends more than 20 answer options; using solution recovery, Green and Rao (1970) suggest researchers should “attempt to secure responses at least at the level of six-point response scales” (p. 38); using inter-rater reliability, Cicchetti, Showalter and Tyrer (1985) recommend seven answer options; and using the correlation with an objective behavioural criterion, Hancock and Klockars (1991) recommend nine-point formats. Miller (1956) recommends approximately seven answer options, based on the argument that this is the number of points the human mind can discriminate; and as a conclusion from his review article, Cox (1980) states that “scales with two or three response alternatives are generally inadequate in that they are incapable of transmitting very much information and they tend to frustrate and stifle respondents.” He recommends “seven plus minus two ” answer options.

Overall, the key belief shared by the proponents of multi-category answer formats is that a low number of answer options does not allow people to differentiate between options sufficiently. This belief is represented well by Garner’s (1960) statement that “it is clear that information transmission cannot be lost by increasing the number of rating categories. Therefore, it is better to err on the side of having too many categories than to err by having too few” (p. 352).

Arguments in support of binary-answer formats

Several researchers come to the exact opposite conclusion; namely that binary-answer formats are preferable or that, at least, how many answer options respondents are offered does not make much difference (Bendig, 1954; Dolnicar & Grün, 2007b, 2007c; Dolnicar, Grün & Leisch, in press; Komorita & Graham, 1965; Matell & Jacoby, 1971a; 1971b; Martin, Fruchter & Mathis, 1974; Schutz & Rucker, 1975). For example, Peabody (1962) concludes that the six-point item format reflects “primarily the direction of responses” (p.73), which is captured equally well by the binary-answer format. He therefore recommends using dichotomous scoring of items. He also concludes that differences in ratings on multi-category items “primarily represent response sets, and only to a secondary degree actual differences in intensity” (p.73). Similarly, Komorita (1963) compared results from a six-point and a binary format, concluding that the correlation between six-point and binary scores is very high, and therefore “Likert’s weighting of item response by intensity had practically no effect on total scores. One may just as well give 0, 1 weights for favourable responses instead of differential weights for intensity and obtain practically the same results” (p. 332).

Key reasons for proponents of the binary format include ease of administration, ease of scoring, avoidance of response styles (e.g., Komorita & Graham, 1965), ease of completion (Jones, 1968), preference by respondents (Dolnicar, 2003), quickness (Dolnicar & Grün, 2007a; 2007c; Preston & Colman, 2000), the fact that too many answer options ask for more discrimination than the respondent is capable of, and, most importantly, that results do not actually provide less information. Sometimes, they argue, the additional dimension of intensity gives a

false sense of more information, capturing additional response sets, rather than true differences in beliefs.

Methodology

Questionnaire

Our questionnaire involved asking respondents to assess evaluative beliefs relating to fast-food brands. Included were five brands (Subway, McDonald's, Red Rooster, KFC, and Pizza Hut), and 11 attributes that emerged from a qualitative pre-study as the key characteristics consumers use to evaluate fast-food restaurants: disgusting, greasy, fattening, fast, expensive, spicy, healthy, tasty, cheap, convenient, and yummy.

The binary-answer format version of the questionnaire asked respondents to evaluate each brand-attribute association with a "yes" if they believed that the brand had the characteristic (e.g., McDonald's is convenient) or a "no" if they believed that the brand did not have the characteristic (e.g., KFC is spicy). This is not the typical way of conducting brand image measurement, and currently the so-called "pick any" measures still dominates brand image research. In the case of "pick any" measures, respondents are not offered both a "yes" and "no" option, instead, for each attribute, only one answer box is offered which has the meaning of "yes." If a respondent does not perceive a brand to have any given attribute, the respondent ticks nothing. Because non-response is an acceptable way of completing the questionnaire, the "pick any" format is prone to evasion error. The seven-point answer format used in our study offered respondents seven answer options, with the endpoints articulating the opposites for each attribute. All options were labeled, as is usually the case with multi-category answer formats used in tourism research.

Preference for each of the fast-food restaurants was measured using the following question: "Please indicate how much you personally like each of the fast-food chains listed below." Respondents were offered a semantic differential answer format with 11 answer options and the endpoints labeled "I love it" and "I hate it." This measure represents an overall attitude, not an attribute or evaluative belief, and is therefore better measured using a single item (Rossiter & Bergkvist, 2009) that offers a numerical answer scale with between five and 11 scale points (Rossiter, 2011).

User-friendliness is assessed by asking respondents two questions and by measuring the actual time it takes them to complete the questionnaire. The two questions about user friendliness were worded as follows: "How did you experience the questionnaire?" (with answer options "easy to answer," "ok," and "difficult") and "How did you feel about completing this questionnaire?" (with answer options "it was fun," "I didn't mind," and "it was annoying").

Fieldwork administration

We conducted a permission-based online survey study to collect the data. Respondents were asked to complete two surveys, one week apart. They were confronted both times with a block of questions that required them to provide evaluative beliefs relating to fast-food restaurants. The final sample used for this analysis consisted of 100 respondents who were offered the binary-answer format and 100 respondents who were presented with the seven-point answer format.

Data analysis

We assessed stability of responses by comparing the answers to all attribute questions across the two survey waves. For each respondent, we calculated the percentage of questions to which the exact same response was given over both waves.

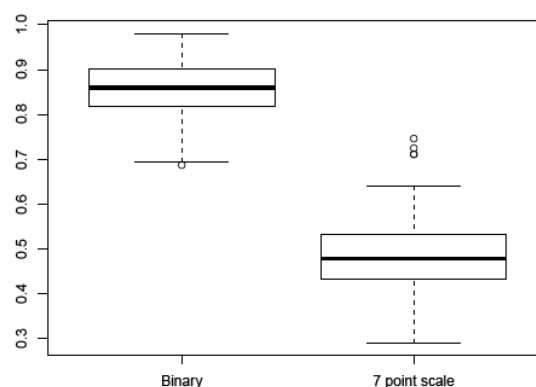
We assessed concurrent validity of responses in the following way: A nonparametric regression model was fitted to the data predicting how much a respondent liked a particular brand (an 11-point scale measuring overall attitude), based only on the evaluative beliefs provided by the respondent. We used random forests (Breiman, 2001) as the regression model, because they can automatically select variables (i.e., perceptions) and can model interactions between them. The cross-validated R^2 value (percentage of variance explained) of the random forest was used as the criterion of evaluation.

All calculations and figures were made using the statistical computing environment R version 2.11.1 (R Development Core Team, 2010) using extension package randomForest (Liaw & Wiener, 2002).

Results

Figure 1 contains the results from the stability comparison. Stability ranges from 0 (zero) to 100 percent and, as shown, the binary-answer format achieves a stability level of 86 percent, thus significantly (t statistic of 34.6, t test p-value < 0.001) outperforming the seven-point answer format, which only reaches a stability level of 48 percent. This result means that Hypothesis 1 (the binary-answer format outperforms the seven-point format with respect to stability over time in the context of measuring evaluative beliefs) is strongly supported by our data.

Figure 1
Stability of brand attribute associations



Regarding concurrent validity using the overall attitude for each one of the rated fast-food chains as the dependent variable, the binary-format again outperforms the seven-point format, with a concurrent validity (cross-validated R^2 of the regression model) of 0.38, compared to 0.06 for the seven-point format.

From the analysis of concurrent validity, we may conclude that the binary format leads to better results than the seven-point format. Consequently, Hypothesis 2 (the binary-answer format leads to higher levels of concurrent validity than the seven-point item format in the context of measuring evaluative beliefs) cannot be rejected.

Table 3 provides results related to the comparison of user-friendliness. As shown, a higher proportion of respondents stated that the binary format was “easy to answer” and “fun,” although neither difference is statistically significant. Regarding the time it took respondents to complete questionnaires using the two alternative answer formats, two respondents needed more than 50 minutes to complete the survey. Because no other respondent required more than 15 minutes, we removed these two outliers from the analysis. After exclusion of these cases, the binary questionnaire took approximately six minutes to complete on average, compared to seven and a half minutes for the seven-point format. Because the observed times have a skew distribution, we conducted a one-sided t-test on log scale for difference in means, resulting in a t-statistic of 3.17 and a p-value of < 0.001. The difference in duration is not only statistically significant; it is substantial enough to be practically meaningful, saving approximately 20 percent of (expensive) fieldwork time.

Table 3
Comparison of duration and user-friendliness

	Binary	Seven point
Easy to answer	86%	81%
OK	14%	18%
Difficult	0%	1%
It was fun	42%	35%
I didn't mind	57%	63%
It was annoying	1%	2%
Average minutes & seconds	06.03	07.28

Based on these results, Hypothesis 3 (the binary-answer format outperforms the seven-point answer format with respect to user-friendliness) cannot be answered either way in the context of measuring evaluative beliefs. Perceived user-friendliness is not significantly different, whereas completion time is. This result is explainable because we assumed that the questionnaire was very short and easy to complete, compared to typical brand image surveys used in market research on a regular basis. Therefore, no matter which answer format was offered, the questionnaire never became tedious. This explanation would have to be tested with new empirical data, including several questions more representative of typical market research studies, in order to be useful to other sectors.

Conclusions

This study examines empirically whether the binary-answer format is indeed inferior to the multi-category answer format that currently dominates as the preferred empirical measure of beliefs in tourism research generally, and evaluative beliefs specifically.

Results indicate that — in the context of measuring evaluative beliefs — the binary-answer format outperforms the seven-point format with respect to stability

over time (H1), and that the binary-answer format leads to higher levels of concurrent validity than the seven-point format (H2). No conclusive result was derived in relation to user-friendliness (H3); while the questionnaire was completed faster by respondents who used the binary scale; there was no significant difference in pleasantness and ease of completion as self-assessed by respondents. We assume that this is because the survey was very short and thus not burdensome, even if respondents used seven-point answer format.

These findings have four major practical implications for academic and applied empirical tourism researchers:

- The current prevailing practice of using certain types of answer formats only because they are the most frequently used in empirical research within the discipline is unacceptable. We call on reviewers and editors to question why the answer formats were chosen in empirical tourism studies in the review process. Invalid measurements lead to invalid conclusions.
- Researchers need to assess — ideally in a small-scale qualitative pre-test — which answer option validly captures responses relating to the construct under study.
- Researchers should provide reasons for their choice of answer format when reporting results.
- Binary-answer formats are unlikely to outperform other answer formats in all contexts. Sometimes, no good theoretical reason exists to believe that binary-answer formats are the better choice (see for example, Rossiter's 2011 justification of numeric, multi-categorical measurement of overall attitude). In the context of measuring evaluative beliefs, the empirical evidence provided in this paper indicates the superior performance of the binary-answer format with two answer options over a seven-point answer format.

This study is limited in three ways. First, we used a strict measure of stability, which could disadvantage the seven-point format because, for example, a change from, a 3 to a 4 is less dramatic than a change from a 0 (zero) to a 1 in the binary format. Therefore, future researchers could conduct a valuable sensitivity analysis to assess the effect of the stability measure on the results. However, the authors maintain a theoretical argument for using the strict measure; namely that we assume that a researcher choosing an answer format believes that each answer option offered to the respondent is actually meaningful to them. If so, they should not be able to reproduce their response when asked twice within a short timeframe.

The second limitation of the study is that this study only compares two answer formats: the binary and the seven-point format. Future work should include a wider range of formats and compare them using the same set of practically relevant criteria: stability, validity, and speed (or user-friendliness). Finally, this study did not attempt to account for scale usage heterogeneity (Rossi, Gilula & Allenby, 2001; DeJong et al., 2008). This major issue needs to be studied in future. The a priori hypothesis proposed by the present study's authors was that the binary-answer format as implemented in here would be less prone to capturing scale usage heterogeneity. If this proves true, it would offer an avenue for eliminating response styles, rather than being forced to correct for them *ex post*, often using questionable algorithms.

References

- Albaum, G., Rogers, R.D., Roster, C. & Yu, J.H. (2006). "Simple rating scale formats exploring extreme response," *International Journal of Marketing Research*, 49(5): 633–650.
- Bendig, A.W. (1954). "Reliability and the number of rating scale categories," *The Journal of Applied Psychology*, 38(1): 38–40.
- Boote, A.S. (1981). "Reliability testing of psychographic scales: five-point or seven-point? Anchored or labelled?" *Journal of Advertising Research*, 21: 53–60.
- Breiman, L. (2001). "Random Forests," *Machine Learning*, 45(1): 5–32.
- Calantone, R., Schewe, C. & Allen, C.T. (1980). "Targeting specific advertising messages at tourist segments," in Hawkins, D.E., Shafer, E.L. & Rovelstad, J. (eds.) *Tourism Marketing and Management*, George Washington University, Washington DC, pp. 133–147.
- Champney, H. & Marshall, H. (1939). "Optimal refinement of the rating scale," *Journal of Applied Psychology*, 23: 323–331.
- Chang, L. (1994). "A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity," *Applied Psychological Measurement*, 18(3): 205–215.
- Cicchetti D.V., Showalter D. & Tyrer, P. (1985). "The effect of number of rating scale categories upon levels of interrater reliability: A Monte Carlo investigation," *Applied Psychological Measurement*, 9: 31–46.
- Cox, E.P. (1980). "The optimal number of response alternatives for a scale: A review," *Journal of Marketing Research*, 17: 407–422.
- Crask, M. (1981). "Segmenting the vacationer market: Identifying the vacation preferences, demographics, and magazine readership of each group," *Journal of Travel Research*, 20: 20–34.
- De Jong, M.G., Steenkamp, J.-B.E.M., Fox, J.-P. & Baumgartner, H. (2008). "Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation," *Journal of Marketing Research*, 45(1): 104–115.
- Dolnicar, S. (2003). "Simplifying three-way questionnaires — Do the advantages of binary answer categories compensate for the loss of information?" in *Australia and New Zealand Marketing Academy (ANZMAC) CD Proceedings*, ANZMAC.
- Dolnicar, S. (2004). "Beyond 'Commonsense Segmentation' — A systematics of segmentation approaches in tourism," *Journal of Travel Research*, 42(3): 244–250.
- Dolnicar, S. & Grün, B. (2007a). "User-friendliness of answer formats — An empirical comparison," *Australasian Journal of Market and Social Research*, 15(1): 19–28.
- Dolnicar, S. & 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. & Grün, B. (2007c). "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., Grün, B. & Leisch, F. (in press). "Quick, simple and reliable: Forced binary survey questions," *International Journal of Market Research*.
- Finn, R.H. (1972). "Effects of some variations in rating scale characteristics on the means and reliabilities of rating," *Educational and Psychological Measurements*, 32: 255–265.
- Garner, W.R. (1960). "Rating scales, discriminatability, and information transmission," *The Psychological Review*, 67(6): 343–352.
- Goodrich, J. (1980). "Benefit segmentation of US international travelers: An empirical study with American Express," in Hawkins, D.E., Shafer, E.L. & Rovelstad, J. (eds.), *Tourism Marketing and Management*, George Washington University, Washington DC, pp. 133–147.
- 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(3): 147–154.
- Jenkins, G.D.J. & Taber, T.D. (1977). "A Monte Carlo study of factors affecting three indices of composite scale reliability," *Journal of Applied Psychology*, 62(4): 392–398.
- Jones, R.R. (1968). "Differences in response consistency and subjects' preferences for three personality inventory response formats," in *Proceedings of the 76th Annual Convention of the American Psychological Association*, pp. 247–248.
- 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 Measurements*, 25: 987–995.
- Liaw, A. & Wiener, M. (2002). "Classification and regression by random forest," *R News*, 2(3): 18–22.
- Lissitz, R.W. & Green, S.B. (1975). "Effect of the number of scale points on reliability: A Monte Carlo approach," *Journal of Applied Psychology*, 60(1): 10–13.
- Martin, W.S.M., 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 Measurements*, 34: 537–545.
- Matell, M.S. & Jacoby, J. (1971a). "Communication and research notes," *Journal of Marketing Research*, 8(4): 495–500.
- Matell, M.S. & Jacoby, J. (1971b). "Is there an optimal number of alternatives for Likert scale items? Study 1: Reliability and validity," *Educational and Psychological Measurements*, 31: 657–674.
- Mazanec, J.A. (1984). "How to detect travel market segments: A clustering approach," *Journal of Travel Research*, 23(1): 17–21.

- Mazanec, J.A. (2000). "Market segmentation," in Jafari, J. (Ed.), *Encyclopedia of Tourism*, Routledge, London.
- Meyer, D., Leisch, F. & Hornik, K. (2003). "The support vector machine under test," *Neurocomputing*, 55: 169–186.
- Miller, G.A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychological Review*, 63): 81–97.
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric Theory* (3rd ed.) McGraw Hill, New York, NY.
- Ooster, T.R.F. (1989). "Number of alternatives per choice point and stability of Likert-type scales," *Perceptual and Motor Skills*, 68: 549–550.
- Peabody, D. (1962). "Two components in bipolar scales: Direction and extremeness," *Psychological Review*, 69(2): 65–73.
- 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 (2010). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Remmers, H.H. & Ewart, E. (1941). "Reliability of multiple-choice measuring instruments as a function of the Spearman-Brown prophecy formula," *Journal of Educational Psychology*, 32(1): 61–66.
- Rossi, P.E., Gilula, Z. & Allenby, G. (2001). "Overcoming Scale Usage Heterogeneity," *Journal of the American Statistical Association*, 96(453): 20–31.
- Rossiter, J.R. (2011). *Measurement for the Social Sciences — The C-OAR-SE Method and Why It Must Replace Psychometrics*, Springer, New York.
- Rossiter, J.R. & Bergkvist, L. (2009). "The importance of choosing one good item for single-item measures of attitude towards the ad and attitude towards the brand and its generalization to all measures," *Transfer: Werbeforschung & Praxis*, 55(2): 8–18.
- Schutz, H.G. & Rucker, M.H. (1975). "A comparison of variable configurations across scale lengths: An empirical study," *Educational and Psychological Measurement*, 35: 319–324.
- Symonds, P.M. (1924). "On the loss of reliability in rating due to coarseness of the scale," *Journal of Experimental Psychology*, 7: 456–461.

Short Bios

Sara Dolnicar completed her PhD at the Vienna University of Economics and Business Administration. She is currently working as a Professor of Marketing at the University of Wollongong in Australia and serves as the Director of the Institute for Innovation in Business and Social Research (IIBSoR). Her research interests include market segmentation, quantitative methodology in marketing research, answer format effects and response styles and tourism marketing.

Friedrich Leisch completed his PhD in Applied Mathematics at the Vienna University of Technology. He is currently working as a Professor of Statistics at the Ludwig-Maximilians-Universität in Munich, Germany. His research interests include statistical computing, cluster analysis, mixture models, statistical learning, and applications in economics, management science and biomedical research.