

2011

Three good reasons NOT to use five and seven point Likert items

Sara Dolnicar

University of Wollongong, sarad@uow.edu.au

Bettina Grun

Wirtschaftsuniversitat Wien, bettina@uow.edu.au

Friedrich Leisch

University of Munich, Germany

John Rossiter

University of Wollongong, jrossite@uow.edu.au

Publication Details

This conference paper was originally published as Dolnicar, S, Grun, B, Leisch, F and Rossiter, J, Three good reasons NOT to use five and seven point Likert items, CAUTHE 2011 : 21st CAUTHE National Conference, Adelaide, Australia, 8-11 February 2011.

Three good reasons NOT to use five and seven point Likert items

Abstract

One of the main sources of knowledge development in tourism has been survey research. Through surveys of tourists, surveys of the tourism industry and surveys of residents living in tourism destinations, our understanding of the positive and negative impact of tourism has increased and we are now better able to understand and respond to tourists' needs. The underlying measure in survey research is the question respondents are asked. In the physical sciences measures are highly calibrated and objective. For example, a thermometer – if not faulty – will show the true temperature of a location at any given time. A ruler will measure the true length of an object. This is not the case in the social sciences. Although, as Nunnally and Bernstein (1994, p. 6) state, the “major advantage of measurement is taking the guesswork out of scientific observation”, in survey research it appears that little attention is paid to the actual measure used. Proof for this statement can be derived from a review of all survey-based research studies published in the *Journal of Travel Research* over the past five years. Thirty-two percent used a seven point answer format and another 33 percent used a five point answer format, most of them not providing a reason for this choice. In the few cases where a reason was provided, it was simply that the answer format was based on previous measures of the construct under study or similar. It appears that the use of five and seven point answer formats has become a habit that, through repetition, has become accepted as the best measure of beliefs. As a consequence, reviewers of tourism journals no longer request a justification for the use of such answer formats. If, however, the authors have good reasons to use an alternative answer format, such as a simple yes-no binary format, reviewers are very sceptical and demand justification at best, or the collection of a new data set at worst. We challenge the currently dominant belief that five and seven point Likert items are the single best measure to capture human beliefs. There are (at least) three good reasons why Likert items in general, and seven point Likert items specifically, can be a dangerous default choice for survey researchers:

1. Seven point Likert items suffer from response style bias. Response styles are response biases which respondents display independent of the content of the questions. Response biases are defined as “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)” (Paulhus, 1991, p.17). It is widely acknowledged that not only do individuals differ in their response styles, but, more critically, groups of respondents coming from different cultures display different response styles (Chun, Campbell and Yoo, 1974; Hui and Triandis, 1989; Marin, Gamba, and Marin, 1992; Marshall and Lee, 1998; Roster, Rogers and Albaum, 2003; van Herk, Poortinga and Verhallen, 2004; Welkenhuysen-Gybels, Billiet and Cambre, 2003; Zax and Takahashi, 1967).
2. Such cross-cultural response styles are a serious problem in tourism research, where often respondents from different cultures are surveyed. Once the data is collected it is hard to differentiate whether, for example, a “7” on the seven point answer format indicates a high level of agreement, or indicates slight agreement plus a response style.

Peabody (1962) concluded that Likert items “primarily represent response sets, and only to a secondary degree actual differences in intensity” (p. 73). He also postulates that Likert items primarily capture direction (positive or negative) and to a much lesser degree intensity (level of agreement or disagreement). It is for this reason that the binary answer format (offering respondents just a “yes” and a “no” option) represents a viable measurement alternative: it captures direction very effectively, but avoids contaminating data with intensity, which is more reflective of response styles than intensity of beliefs. Peabody, along with other authors (e.g.

Komorita and Graham, 1965), recommended the use of binary answer formats to avoid response styles many decades ago. But, unfortunately, their voices were not heard by the scientific community.

1. Seven point Likert items are less stable than binary answer formats. This is demonstrated by our work, for which we have collected the data and conducted preliminary analysis. The full results will be presented at this conference.

Bendig (1954), Komorita and Graham (1965), Matell and Jacoby (1971a and 1971b), Martin, Fruchter and Martin (1974) and Schutz and Rucker (1975) all conclude from their (mostly empirical studies) that binary answers perform as well as multi-category answer formats on all criteria.

1. Both five and seven point Likert items take longer to complete. In studies comparing the time it takes respondents to complete the questionnaire using different answer formats, Jones (1968), Preston and Colman (2000), Dolnicar (2003) and Dolnicar and Grün (2007a and b) conclude that the binary answer format was completed faster than multi-category answer formats, supporting a finding by Preston and Coleman (2002). A number of previous authors have come to similar conclusions. For example, Komorita and Graham (1965) argued that the binary answer format is easier to administer and easier to score. Jones (1968) found that binary answer formats were easier for respondents to complete.

From the work of various pioneers in the area of questionnaire design and our own work, a number of key conclusions can be drawn which have very direct and practical implications for survey researchers in tourism, both in industry and in academia:

- First and foremost: a survey question is a scientific measure and must be treated as such. Its format has to be thought through, justified and pretested before it can be assumed to be a valid measure of the construct under study. The results of this thinking and pretesting process need to be made available to readers of academic articles as well as market research reports.
- Secondly, there is no good explanation for the current dominance of the five and seven point Likert items. While valuable in some instances, simpler, quicker, easier and more valid measures could be used without loss of predictive validity.

Disciplines

Business | Social and Behavioral Sciences

Publication Details

This conference paper was originally published as Dolnicar, S, Grun, B, Leisch, F and Rossiter, J, Three good reasons NOT to use five and seven point Likert items, CAUTHE 2011 : 21st CAUTHE National Conference, Adelaide, Australia, 8-11 February 2011.

Three good reasons NOT to use five and seven point Likert items

Sara Dolnicar^{a*}, Bettina Grün^c, Friedrich Leisch^b and John Rossiter^a

^a Institute for Innovation in Business and Social Research
Faculty of Commerce, University of Wollongong,
Wollongong, NSW 2522, Australia
Telephone: (61 2) 4221 3862
sara_dolnicar@uow.edu.au, john_rossiter@uow.edu.au

^b Department of Statistics
Ludwig-Maximilians-Universität München
Ludwigstrasse 33, 80539 München, Germany
Telephone (49 89) 2180 3165
Friedrich.Leisch@stat.uni-muenchen.de

^c Institute for Statistics and Mathematics
Wirtschaftsuniversität Wien, Augasse 2-6, 1090 Wien, Austria
Telephone (43 1) 31336 5032
Bettina.Gruen@wu.ac.at

*Authors listed in alphabetical order

Working Paper

One of the main sources of knowledge development in tourism has been survey research. Through surveys of tourists, surveys of the tourism industry and surveys of residents living in tourism destinations, our understanding of the positive and negative impact of tourism has increased and we are now better able to understand and respond to tourists' needs.

The underlying measure in survey research is the question respondents are asked. In the physical sciences measures are highly calibrated and objective. For example, a thermometer – if not faulty – will show the true temperature of a location at any given time. A ruler will measure the true length of an object. This is not the case in the social sciences. Although, as Nunnally and Bernstein (1994, p. 6) state, the “major advantage of measurement is taking the guesswork out of scientific observation”, in survey research it appears that little attention is paid to the actual measure used.

Proof for this statement can be derived from a review of all survey-based research studies published in the Journal of Travel Research over the past five years. Thirty-two percent used a seven point answer format and another 33 percent used a five point answer format, most of them not providing a reason for this choice. In the few cases where a reason was provided, it was simply that the answer format was based on previous measures of the construct under study or similar.

It appears that the use of five and seven point answer formats has become a habit that, through repetition, has become accepted as the best measure of beliefs. As a consequence, reviewers of tourism journals no longer request a justification for the use of such answer formats. If, however, the authors have good reasons to use an alternative answer format, such as a simple yes-no binary format, reviewers are very sceptical and demand justification at best, or the collection of a new data set at worst.

We challenge the currently dominant belief that five and seven point Likert items are the single best measure to capture human beliefs. There are (at least) three good reasons why Likert items in general, and seven point Likert items specifically, can be a dangerous default choice for survey researchers:

1. Seven point Likert items suffer from response style bias. Response styles are response biases which respondents display independent of the content of the questions. Response biases are defined as “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)” (Paulhus, 1991, p.17). It is widely acknowledged that not only do individuals differ in their response styles, but, more critically, groups of respondents coming from different cultures display different response styles (Chun, Campbell and Yoo, 1974; Hui and Triandis, 1989; Marin, Gamba, and Marin, 1992; Marshall and Lee, 1998; Roster, Rogers and Albaum, 2003; van Herk, Poortinga and Verhallen, 2004; Welkenhuysen-Gybels, Billiet and Cambre, 2003; Zax and Takahashi, 1967).

2. Such cross-cultural response styles are a serious problem in tourism research, where often respondents from different cultures are surveyed. Once the data is collected it is hard to differentiate whether, for example, a “7” on the seven point answer format indicates a high level of agreement, or indicates slight agreement plus a response style.

Peabody (1962) concluded that Likert items “primarily represent response sets, and only to a secondary degree actual differences in intensity” (p. 73). He also postulates that Likert items primarily capture direction (positive or negative) and to a much lesser degree intensity (level of agreement or disagreement). It is for this reason that the binary answer format (offering respondents just a “yes” and a “no” option) represents a viable measurement alternative: it captures direction very effectively, but avoids contaminating data with intensity, which is more reflective of response styles than intensity of beliefs.

Peabody, along with other authors (e.g. Komorita and Graham, 1965), recommended the use of binary answer formats to avoid response styles many decades ago. But, unfortunately, their voices were not heard by the scientific community.

3. Seven point Likert items are less stable than binary answer formats. This is demonstrated by our work, for which we have collected the data and conducted preliminary analysis. The full results will be presented at this conference.

Bendig (1954), Komorita and Graham (1965), Matell and Jacoby (1971a and 1971b), Martin, Fruchter and Martin (1974) and Schutz and Rucker (1975) all conclude from their (mostly empirical studies) that binary answers perform as well as multi-category answer formats on all criteria.

4. Both five and seven point Likert items take longer to complete. In studies comparing the time it takes respondents to complete the questionnaire using different answer

formats, Jones (1968), Preston and Colman (2000), Dolnicar (2003) and Dolnicar and Grün (2007a and b) conclude that the binary answer format was completed faster than multi-category answer formats, supporting a finding by Preston and Coleman (2002). A number of previous authors have come to similar conclusions. For example, Komorita and Graham (1965) argued that the binary answer format is easier to administer and easier to score. Jones (1968) found that binary answer formats were easier for respondents to complete.

From the work of various pioneers in the area of questionnaire design and our own work, a number of key conclusions can be drawn which have very direct and practical implications for survey researchers in tourism, both in industry and in academia:

- First and foremost: a survey question is a scientific measure and must be treated as such. Its format has to be thought through, justified and pretested before it can be assumed to be a valid measure of the construct under study. The results of this thinking and pretesting process need to be made available to readers of academic articles as well as market research reports.
- Secondly, there is no good explanation for the current dominance of the five and seven point Likert items. While valuable in some instances, simpler, quicker, easier and more valid measures could be used without loss of predictive validity.

Acknowledgements

This research has been funded by the Australian Research Council through the ARC Discovery (DP0878423) and Linkage International (LX0881890 and LX0559628) Grant Schemes and the Austrian Science Fund through the FWF Hertha-Firnberg Grant Scheme (T351).

References

- Bendig, A.W. (1954). Reliability and the Number of Rating Scale Categories. *The Journal of Applied Psychology*, 38(1), 38-40.
- Chun, K.T., Campbell, J.B., & Yoo, J.H. (1974). Extreme response style in cross-cultural research – Reminder. *Journal of Cross-Cultural Psychology*, 5(4), 465-480.
- Dolnicar, S., (2003). Simplifying Three-way Questionnaires - Do the Advantages of Binary Answer Categories Compensate for the Loss of Information? *Australia and New Zealand Marketing Academy (ANZMAC) CD Proceedings*.
- Dolnicar, S., & Grün, B. (2007a). User-Friendliness of Answer Formats – An Empirical Comparison. *Australasian Journal of Market & Social Research*, 15(1), 19-28.
- Dolnicar, S., & Grün, B. (2007b). How Constrained a Response: A Comparison of Binary, Ordinal and Metric Answer Formats. *Journal of Retailing and Consumer Services*, 14(2), 108-122.
- Hui, C.H., & Triandis, H.C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20(3), 296-309.
- 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.

- Komorita, S.S., & Graham, W.K. (1965). Number ID Scale Points and the Reliability Scales. *Educational and Psychological Measurements*, 25, 987-995.
- Marin, G., Gamba, R.J., & Marin, B.V. (1992). Extreme response style and acquiescence among Hispanics - The role of acculturation and education. *Journal of Cross-Cultural Psychology*, 23(4), 498-509.
- Marshall, R., & Lee, C. (1998). A cross-cultural, between-gender study of extreme response style. In *European Advances in Consumer Research Vol. 3* (pp. 90-95). The Association for Consumer Research.
- Martin, W.S.M., Fruchter, B., & Matis, W.J. (1974). An Investigation of the effect of the number of scale intervals on principal components factor analysis. *Educational and Psychological Measurements*, 34(3), 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.
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric Theory*. New York: McGraw-Hill.
- Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver & L.S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp.17-59). San Diego: Academic Press.
- 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.
- Roster, C. A., Rogers, R., & Albaum, G. (2003) A cross-cultural use of extreme categories for rating scales. In S.M. Smith (Ed.), *Proceedings of Ninth Annual Cross Cultural Research Conference*.
- 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.
- van Herk, H., Poortinga, Y.H., & Verhallen, T.M.M. (2004). Response styles in rating scales - Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, 35(3), 346-360.
- Welkenhuysen-Gybels, J., Billiet, J., & Cambre, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology*, 34(6), 702-722.
- Zax, M., & Takahashi, S. (1967). Cultural influences on response style: Comparison of Japanese and American college students. *Journal of Social Psychology*, 71, 2-10.