1-1-2006

Are we drawing the right conclusions? The Dangers of Response Sets and Scale Assumptions in Empirical Tourism Research

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

Recommended Citation

Are we drawing the right conclusions? The Dangers of Response Sets and Scale Assumptions in Empirical Tourism Research

Abstract

Empirical tourism research has a long history and empirically based findings represent an important component of theory development and managerial insight. Nevertheless, empirical data of any kind is susceptible to misinterpretation. The aim of this study is to investigate to which extent empirical tourism research accounts for three sources of potential misinterpretation of results: (1) the occurrence of answer format effects, (2) the occurrence of culturally specific response styles, and (3) the selection of data analytic techniques appropriate for the data format. A review of 43 academic publications from 2000 and 2001 suggests that empirical tourism research is strongly guided by standards which have developed within the tourism research community and are not questioned anymore: ordinal answer formats dominate the field, ordinal data is analyzed using techniques requiring metric data and cross-cultural response styles are ignored, which is a particularly concerning finding given the amount of cross-cultural comparisons typically undertaken in tourism research. Recommendations for improvement are made.

Keywords

answer format effects, response styles, cross-cultural research

Disciplines

Business | Social and Behavioral Sciences

Publication Details

Are we drawing the right conclusions?
The dangers of answer format effects in empirical tourism research

Sara Dolnicar, University of Wollongong

Send correspondence to Sara Dolnicar, School of Management & Marketing, marketing research innovation centre (mric), University of Wollongong, Northfields Ave, 2522 Wollongong, NSW, Australia; Tel: ++61 (2) 4221 3862; Fax: ++61 (2) 4221 4154 (Email: sara_dolnicar@uow.edu.au)

ABSTRACT

Empirical tourism research has a long history and empirically based findings represent an important component of theory development and managerial insight. Nevertheless, empirical data of any kind is susceptible to misinterpretation. The aim of this study is to investigate to which extent empirical tourism research accounts for three sources of potential misinterpretation of results: (1) the occurrence of answer format effects, (2) the occurrence of culturally specific response styles, and (3) the selection of data analytic techniques appropriate for the data format. A review of 43 academic publications from 2000 and 2001 suggests that empirical tourism research is strongly guided by standards which have developed within the tourism research community and are not questioned anymore: ordinal answer formats dominate the field, ordinal data is analyzed using techniques requiring metric data and cross-cultural response styles are ignored, which is a particularly concerning finding given the amount of cross-cultural comparisons typically undertaken in tourism research. Recommendations for improvement are made.

Key words: answer format effects, response styles, cross-cultural research

INTRODUCTION

Imagine that a tourism researcher asks 500 Greek and 500 Korean tourists visiting a destination to complete a questionnaire. The questionnaire includes statements about motivations to visit this particular destination using a five point Likert scale. After the data is collected, this researcher investigates differences between the motivations of Greek and Korean tourists by computing an analysis of variance. The results show that the Greek tourists state to be motivated by many more of the listed benefits than this is the case for Korean travelers. These findings are passed on to the destination's National Tourism Organization (NTO), which consequently invests substantial amounts of money communicating benefits to Greek customers; quite likely without an effect on demand among Greek tourists.

By drawing such conclusions from the research design described above, the researcher could have made at least three major mistakes:

- If Greek respondents tend to use more extreme values on the answer scale or tend to express agreement more frequently than Korean respondents (independent of the question asked), the researcher would have claimed higher motivation levels although the results merely reflect cross-cultural response styles. The presence of people from various countries, continents and cultural backgrounds is in the very nature of tourism. Consequently, tourism research frequently involves respondents who are heterogeneous with regard to their cultural background. Tourism researchers have to be aware of the possible distortions of results.
- If, within both the Greek and Korean tourists, the selected answer format impacts on the answers, the results would be a mixture of response style effects and opinions expressed by respondents.
- And finally, if the five point Likert Scale is not demonstrated to have metric data properties on item level, the analysis of variance would have violated the technique's data assumptions, potentially leading to incorrect conclusions about the significance of differences.

Quite possibly, if the researcher had corrected for inter- and cross-cultural response styles and used a technique suitable for ordinal data, no differences between Greek and Korean tourists would have occurred, saving the NTO advertising expenditures with a questionable return on investment.

The aim of this paper is to investigate, whether the example described above is a rare exception in empirical tourism research or not. A review of quantitative empirical tourism research is conducted to determine if the three dangers mentioned above are satisfactorily accounted for in tourism research: (1) the distortion of results due to answer format effects, (2) the distortion of results due to cross-cultural response styles, and (3) the violation of assumptions of data analytic procedures.

The relevance of the article lies in preventing wrong conclusions to be drawn from empirical tourism research, which is of central practical importance to academics, tourism industry, tourism authorities and tourism market research companies using empirical data as basis for their research and subsequently their marketing activities.

PRIOR WORK

Answer format effects

How to best ask questions to get valid results is of fundamental importance to research of all disciplines using empirical data. Consequently, a vast body of literature investigating questionnaire design effects exists in psychology, sociology, psychometrics as well as marketing. One subset of work in this area is of particular interest for the undertaken review of tourism studies: research on effects of different response formats in questionnaires. Typically, these studies aim to determine which response format is optimal, where optimality is defined differently in different studies.

A number of authors choose to define optimality as highest level of reliability. The majority of these studies conclude that the number of points on the answer format given to respondents does not influence reliability (Bendig 1954; Peabody 1962; Komorita 1963; Komorita and Graham 1965; Matell and Jacoby 1971; Jacoby and Matell 1971; Remington, Tyrer, Newson-Smith and Cicchetti 1979; Preston and Colman 2000). A few studies, however, draw the opposite conclusion, recommending the use of higher numbers of scale points (Symonds 1924; Nunnally 1967; Oaster 1989; Finn 1972; Ramsay 1973).

The studies using high validity levels as the criterion for optimality, lead to similar findings. Matell and Jacoby (1971), Jacoby and Matell (1971), Chang (1994) and Preston and Colman (2000) conclude that the choice of answer format is not associated with different levels of validity. Contrarily, the results obtained by Loken, Pirie, Virnig, Hinkle and Salmon (1987) and Hancock and Klockars (1991) indicate that a larger number of options (for instance using a seven point scale as opposed to a five point scale) increases validity.

A third stream of research uses structural equivalence as criterion to evaluate answer format effects. Typically, factor analyses are computed for different answer formats and the results compared. Martin, Fruchter and Mathis (1974), Percy (1976) and Green and Rao (1970) have chosen this research approach. Green and Rao recommend using at least six answer options as a consequence of their study. Percy found no significant differences in the underlying factors resulting from different answer formats.

Finally, a few authors have investigated the perspective of consumer-friendliness of surveys. Jones (1968) and Preston and Colman (2000) come to the conclusion that respondents prefer having the choice between more options. Dolnicar (2003) and Dolnicar, Grun and Leisch (2004) conclude that ordinal scales are perceived as significantly more difficult to answer and take significantly more time to complete.

No optimal answer exists to the question of how many points on an answer format are best. While some authors recommend using binary (yes-no) options (Peabody, 1962; Matell and Jacoby, 1971; Jacoby and Matell, 1971; Komorita and Graham, 1965; Dolnicar, Grun and Leisch, 2004), a frequency count of recommendations across studies on answer formats would lead to the use of seven point scales according to Cox (1980).

The popularity of such ordinal scales in marketing research in general has been noted by Peterson (1997), Van der Eijk (2001) and Dolnicar (2002). The methodological dangers of both ordinal and binary scales have been extensively discussed by Scharf (1991), Peterson (1997) and Kampen and Swyngedouw (2000) who highlight in particular the difficulties of assuming equidistance between ordinal scale categories and data analytic consequences thereof. For instance, five point Likert scales are typically treated as interval scale level data, which was never intended by Likert (1932) who claimed metric properties

only for the summated scale not the single items. Kampen and Swyngedouw (2000) distinguish different kinds of ordinal scales and conclude that ordinal scales representing unstandardized discrete variables with ordered categories, such as the agreement scales typically used in tourism research, have very undesirable properties which can best be described as follows (p. 99) "in many instances the experimenter can only hope that in general respondents or experimentators attach the same meaning to the categories of an ordinal variable." The reason is that no objective standard defining the scale points exists and significant calibration would be required to get all respondents to have the same understanding of each scale point.

Billiet and McClendon (2000), McClendon (1991) and Watson (1992) draw attention to another problem inherent in Likert scales: the susceptibility to acquiescence or yeasaying bias.

Empirical tourism researchers have to choose which answer format is best suited for their research problem with respect to the advantages and disadvantages of different formats: (1) the quality of the data, (2) the susceptibility to response styles, (3) the amount of time required answering questions, and (4) the analytic techniques planned to be used when data is available.

Cross-cultural response styles

Tourism research typically includes individuals from different cultural backgrounds. In many cases, empirical research in tourism aims at revealing differences between cultural groups or tourists from different countries of origin and is thus exposed to a number of potential distortions of results such as culturally biased response norms which can cause different scale usages independent of the questions; or culturally different interpretations of questions; or differences in the underlying constructs measured. The most concerning potential mistake that results from cross-cultural response styles is that differences in groups means are no longer interpretable due to response styles (Chun, Campell and Yoo, 1974).

The issue of potential dangers of interpreting empirical data derived from surveys conducted in different languages in different places has not been broadly discussed within the tourism research literature so far. Prior work in the area is consequently drawn from other disciplines, such as psychology, sociology and market research, where cross-cultural issues in empirical research have been extensively investigated.

The main foci of these studies are twofold: a series of investigations aimed at revealing whether socio-demographic variables are associated with response styles (Arce-Ferrer and Ketterer, 2003; Albaum, 1997; Bachman and O'Malley, 1984; Bryne and Campbell, 1999; Chun, Campell and Yoo, 1974; Greenleaf, 1992; Hui and Triandis, 1989; Marin, Gamba, Marin, 1992; Van Herk, Poortinga and Verhallen, 2004). The major findings that have repeatedly emerged from these studies are that African-American, Hispanic respondents as well as respondents from numerous countries located at the Mediterranean tend to give more extreme answers. Contrarily, Asian respondents tend to use the extreme options less. Furthermore, higher age, lower education levels and lower household income levels have been found to be significantly associated with extreme response styles.

The second stream of research is methodological in nature and introduces techniques to determine the existence of response styles in cross-cultural studies and correct for the bias (Cheung and Rensvold, 2000; Byrne and Campbell, 1999; Greenleaf,

1992a and 1992b; Van de Vijver and Poortinga, 2002; Welkenhuysen-Gybels, Billiet and Cambre, 2003) with recommendations ranging from very simple approaches, such as investigating if systematic response patterns can be detected for the same cultural group, to modeling approaches to extract the extreme response and acquiescence bias from the actual information content and then correct the data accordingly.

Strong empirical support has been presented in the past to support that different cultures have different response styles as such findings have been replicated repeatedly in different contexts over the past three decades; these findings are highly relevant to empirical tourism research. If the level of distortion of results through cross-cultural response styles is not evaluated for empirical data sets before analysis is undertaken, the probability is very high that wrong conclusions will be drawn by confounding response styles with actual answers.

METHOD

The method selected to investigate the research aims of this study is a review of academic tourism research published in the Annals of Tourism Research or Tourism Management in the years of 2000 and 2001. All articles were screened from the respective volumes and included articles classified as quantitative empirical studies of tourists. This means that empirical work centering on the tourism industry and residents at tourism destinations as respondents were excluded, even if quantitative and empirical in nature. Furthermore, data had to be available in disaggregate form: analyses of aggregate statistics such as national tourism statistics or demand time series were excluded. Finally, purely descriptive studies were excluded, in which authors reported only means or frequency counts of socio-demographics.

The final data set consists of 43 studies (see Appendix for references). Each one of these studies was coded along 51 variables, some of which were simple identifiers as author name, year and journal. The majority of variables, however, code the format in which data was collected, whether cross-cultural aspects were involved and in which way specific data formats were analyzed. Coding was undertaken separately for each of the main constructs investigated in the published studies.

Descriptive statistics are computed to analyze the data set using SPSS in its 12.1.0 version.

RESULTS

First, the article describes general descriptive information on the empirical tourism studies investigated, and then the main analysis regarding the use of answer formats is presented. Among the constructs studied, tourist satisfaction emerges as the most frequently researched topic, followed closely by perceptions and attitudes. Table 1 provides the numbers of studies investigating each of the constructs. The absolute numbers of studies do not add up to 43, as could be expected. The reason is that many studies investigate more than one construct. In such cases, all constructs in the data set are coded.

Despite the popularity of perception and satisfaction studies, a wide variety of other constructs are studied: half of all coded constructs were studied three times or less. This indicates the broadness of constructs investigated empirically in tourism research.

---------- insert Table 1 here ----------

On average, the data sets used by tourism researchers included 1169 respondents, with the smallest one containing 40 and the largest one 8842 respondents. The average response rate was 56 percent, ranging from a low of 17 to a high of 92. Only 15 studies included information on the response rate achieved. Follow-up contacting is not common among tourism researchers: only seven percent use this tool to increase response rates and reduce response bias. Clearly, one reason for low follow-up rates is the high proportion of cross-sectional studies questioning respondents during their stay, making follow-ups very difficult.

Another interesting observation applies to the age of data sets. Among the studies that stated when the data were collected, the average age of the data set analyzed is 4 years, with the most recent being 2 years old and the oldest 11. Given that none of the data sets was used to merely demonstrate a technique, data age issues are concerning as market dynamics might have changed dramatically since the collection of data.

With respect to the answer formats typically used by tourism researchers, Table 2 shows that ordinal scales dominate empirical tourism research, which provides the chosen answer format separately for constructs measured (in the first 7 rows) as well as the answer format chosen for the main construct of the study, where more than one construct was investigated in one publication. Columns 2-6 give the absolute number of studies where certain answer formats were chosen for particular constructs, column 7 gives the percentage of studies within each construct that used ordinal scales of some form.

As can be seen, one construct exists that is not typically investigated with ordinal scales: behavior. This finding is reasonable as questions about past or present behavior can easily be formulated in binary format without concerns that any relevant information might be lost. This same reasoning does not seem to apply when behavioral intentions are measured. Across all constructs, three quarters of the studies use ordinal scales. Among users of ordinal scales, the five point scale is most popular with 40 percent of authors using it, followed by 16 percent using 7 scale points, 13 percent using nine and the remaining researchers choosing less than 5 options.

--------- insert Table 2 here ----------

Given the difficulties associated with the use of ordinal scales, investigating the authors' reasoning for the predominant use of ordinal scales is of particular interest. The explanations provided by authors were coded in the data. Table 3 includes the results, indicating that the vast majority - almost two thirds of authors - does not justify their choice of answer format. Fourteen percent do not need to explain their choice because the measurement scale is a logical consequence of the construct measured. For instance, if respondents are asked to state their expenditures, dollar amounts on a metric scale are the natural choice. One in ten authors cite prior work that used similar answer formats to investigate the same construct; seven percent use pre-existing scales for the constructs investigated; five percent cite another author who explained why the use of such an answer format is preferable; and only two percent actually state the reasons for the chosen answer format in the manuscript.

---------- insert Table 3 here ----------

Table 4 includes the results on data analytic techniques used by researchers based on the data sets discussed above. The values given in the columns of Table 4 are the absolute number of studies applying each of the techniques listed in the first column given certain data formats. The cells shaded in gray show violations of assumptions unless non-violation is investigated and shown not to have occurred. Again, standards have apparently emerged which are not questioned anymore. In case of the selection of data analytic techniques this phenomenon is, however, even more surprising than this is the case in the context of answer formats. No clear answers can be given which answer format is best, but for all of the data analytic procedures in Table 4 the assumptions and data requirements are generally known. The typical justification for applying techniques requiring metric data to ordinal data is the assumptions that ordinal scales have interval properties, which Likert (1932) does indeed state for the summated scale, but not for the individual items. However, summated scales are not typically used in empirical tourism research, where profiles and differences on item-level are of primary interest. Interestingly even this typical argument was not stated by authors of the studies included in this review.

---------- insert Table 4 here ----------

Next the frequency of samples containing respondents from more than one cultural background is discussed. National Tourism Organizations typically compare guest survey results across countries of origin to determine differences between tourists of different nationalities. Knowing that strong cultural response effects have been shown repeatedly to exist, the question arises whether this serious danger of misinterpretation of results is accounted for in empirical tourism research. Table 5 shows how many studies included respondents from different countries or cultural backgrounds and how many investigated potential distortions that could arise from cross-cultural response styles being misinterpreted as differences in the construct investigated. Only 41 studies were included because two did not give enough information on the sample to understand whether cross-cultural effects could be problematic.

---------- insert Table 5 here ----------

As can be seen, more than half (53 percent) of the empirical tourism studies reviewed did indeed include respondents from different countries. For these data sets respondents from different countries are likely to react to certain answer formats in different ways, independent of the questions asked and the underlying constructs. None of the studies mentioned the problem of possible cross-cultural differences in response styles and discussed the implications for their findings. Two studies, however, were classified as running only low risk of misinterpretation. These studies used expenditure data. To the author's knowledge no studies have so far demonstrated cross-cultural response effects in the context of stating expenditures. All other studies (91 percent of all investigations including respondents from different cultural backgrounds) run a high danger of misinterpreting results.

CONCLUSIONS

A bad selection of the answer format, the measurement tool in empirical tourism research, not only impacts on quality of the results in general, systematic data distortions may result which, if not questioned, can cause wrong conclusions. The tourism researcher is responsible to thoroughly and conscientiously evaluate alternative answer formats for a given construct to be measured, to consider their advantages and disadvantages for the construct under study, to select the most appropriate format and explain the choice to the reader.

Unfortunately, the review of empirical tourism studies reveals that virtually no discussion of answer format choice among empirical tourism researchers has taken place. On the contrary, standards have apparently emerged with respect to answer format choice which are justified by citing past tourism researchers' work, but are not questioned anymore. This is a concerning finding given the proven influence of the chosen answer format on the results.

Data analysis conducted by the reviewed articles demonstrates a high level of violation of data assumptions of techniques: 56 percent of all the methods applied to all data sets were in violation of the assumptions underlying the techniques used. This does not mean that the findings would necessarily be different if appropriate techniques had been used, but potentially this could be the case. Tourism researchers should match the chosen answer format with an analytic technique developed for this particular data format.

Finally, cross-cultural response styles are generally neglected among tourism researchers, although about half of the studies reviewed were based on samples including respondents from different cultural backgrounds. The majority of these studies used ordinal answer formats and was consequently at high risk of misinterpreting results.

In summary, unfortunately, many decades of questionnaire design research has not resulted in the single only perfect answer format. An empirical researcher can therefore legitimate choose whichever scale she or he has evaluated to best measure the construct investigated. But the process of selection should be thorough, clearly explained and the consequences for data analysis have to be actively dealt with.

The fact that response styles are widely ignored in empirical tourism research can potentially lead to dramatic misinterpretations of results. Answer tendencies can be so strong that managerial recommendations like "the Spanish tourists are much more interested in culture; tourism authorities should thus focus on advertising cultural attractions to this group" could be factually incorrect and only an artifact of respondents' Spanish heritage, which is known to favor extreme points of the answer format.

The literature includes many approaches for checking how strongly the data is affected by response styles. Possible techniques include the investigation of the invariance of factorial structures (Van de Vijver and Poortinga, 2002; Van Herk, Poortinga and Verhallen, 2004) underlying the responses of different (cultural) subgroups in order to investigate structural equivalence of the constructs measured as well as investigations of item response patterns across all questions in the questionnaire (Byrne and Campbell, 1999). If the number of countries in the data set is low, Van de Vijver and Poortinga (2002) recommend regression procedures to investigate the associations of constructs with context variables. Greenleaf (1992a and b) suggests the selection of a specific subset of extreme response style variables that can subsequently be used to determine – across a variety of

constructs – the information and bias proportions of answers, which enables corrections for the distortions before data analysis.

Another option is to choose answer formats which are less susceptible to response styles, such as binary or metric format. Cronbach (1946; 1950) recommends such answer formats to reduce the danger of contamination of data with cross-cultural response tendencies. Similarly Hui and Trinadis (1989) show how the selection of answer format impacts on the occurrence of response styles; they find that the ten point scale weakened extreme response style effects which were identified on a five point scale to be very strong between Hispanic and non-Hispanic respondents.

Whichever answer format is chosen, researchers should investigate whether certain subgroups of the sample demonstrate systematic deviations from the sample and critically question if this could be a result of cross-culturally different response behavior before making too strong claims about the results (Chun, Campell and Yoo, 1974).

Tourism researchers have to be aware of the potential dangers of interpreting empirical data uncritically. They should make informed decisions about the answer format in view of data quality, speed of data collection, simplicity for the respondent, low susceptibility to response styles and data analytic procedures to be used. This decision and the underlying reasoning should be clearly explained to the reader, as should the match of the data analytic technique to the data properties and the way in which possible data distortions were investigated and corrected, if necessary, before proceeding to analyze and interpret data.

References

Albaum, G. (1997). The Likert Scale Revisited: An Alternate Version. *Journal of the Market Research Society, 39*(2), 331-348.

Arce-Ferrer, A. J., & Ketterer, J. J. (63). 3. *Educational and Psychological Measurement, 63*(3), 484-501.

Bendig, A.W. (1954). Reliability and the number of rating scale categories, *Journal of Applied Psychology*, 38 (1), 38-40.

Byrne, B. M., & Campbell, T. L. (1999). Cross-Cultural Comparisons and the Presumption of Equivalent Measurement and Theoretical Structure - A Look Beneath the Surface. *Journal of Cross-Cultural Psychology, 30*(5), 555-574.

Chang, L. (1994). A psychometric evaluation of four-point and six-point Likert-type scales in relation to reliability and validity, *Applied Psychological Measurement*, 18, 205-215.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research using Structural Equation Modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187-212.

Chun, K. T., Campbell, J. B., and J.H. Yoo (1974). Extreme Response Style in Cross-Cultural Research – Reminder, *Journal of Cross-Cultural Psychology*, 5(4), 465-480.

Cox, Eli P. (1980). The optimal number of response alternatives for a scale: A review, *Journal of Marketing Research*, 17 (4), 407-422.

Cronbach, L. J. (1946). Response Sets and Test Validity. *Educational and Psychological Measurement, 6*, 475-494.

Cronbach, L. J. (1950). Further evidence on response sets and test design, *Educational and Psychological Measurement*, 10, 3-31.

Dolnicar, S. (2002). Review of Data-Driven Market Segmentation in Tourism, *Journal of Travel & Tourism Marketing*, 12(1), 1-22.

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

Finn, R. H. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings, *Educational and Psychological Measurement*, 32 (2), 255-265.

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

Greenleaf, E. A. (1992a). Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *Journal of Marketing Research, 29*(May), 176-188.

Greenleaf, E. A. (1992b). Measuring Extreme Response Style. *Public Opinion Quarterly, 56*(3), 328-351.

Hancock, G. R. and A. J. Klockars (1991). The effect of scale manipulations on validity: Targeting frequency rating scales for anticipated performance levels, *Applied Ergonomics*, 22 (3), 147-154.

Hui, C. H., and H.C. Triandis (1989). Effects of Culture and Response Format on Extreme Response Style, *Journal of Cross-Cultural Psychology*, 20(3), 296-309.

Jacoby, Jacob and Michael S. Matell (1971). Three-Point Likert Scales Are Good Enough, *Journal of Marketing Research*, 8, 495-500.

Jones, Richard 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, Jarl and Marc Swyngedouw (2000). The multi-category controversy revisited, *Quality & Quantity*, 34 (1), 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. and W.K. Graham (1965). Number of scale points and the reliability of scales, *Educational and Psychological Measurement*, 25(4), 987-995.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology, 140*, 44-53.

Loken, Barbara, Phyllis Pirie, Karen A. Virnig, Ronald L. Hinkle, and Charles T. Salmon (1987). The Use of 0-10 Scales in Telephone Surveys, *Journal of the Market Research Society*, 29 (3), 353-362.

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.

Martin, Warren S., Benjamin Fruchter and William J. Mathis (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, Michael S. and Jacob Jacoby (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity, *Educational and Psychological Measurement*, 31, 657-74.

Nunnally, Jum C. (1967). *Psychometric Theory*. New York: McGraw-Hill, 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.

Peabody, Dean (1962). Two components in bipolar scales: direction and extremeness, *Psychological Review*, 69 (2), 65-73.

Peabody, Dean (1962). Two components in bipolar scales: direction and extremeness, *Psychological Review*, 69 (2), 65-73.

Percy, Larry (1976). An Argument in Support of Ordinary Factor Analysis of Dichotomous Variables, in *Advances in Consumer Research* Vol. III.

Preston, Carolyn C. and Andrew M. Colman (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences, *Acta Psychologica*, 104, 1-15.

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

Scharf, Andreas (1991). *Konkurrierende Produkte aus Konsumentensicht*. Frankfurt: Verlag Harri Deutsch.

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

van de Vijver, F. J. R., & Poortinga, Y. H. (2002). Structural Equivalence in Multilevel Research. *Journal of Cross-Cultural Psychology, 33*(2), 141-156.

van der Eijk, Cees (2001). Measuring agreement in ordered rating scales, *Quality & Quantity*, 35 (3), 325-41.

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.

Appendix 1: Reviewed literature

Asplet, M., and Cooper, M. (2000). Cultural designs in New Zealand souvenir clothing: the question of authenticity. *Tourism Management, 21*(3), 307-312.

Baker, D. A., and Crompton, J. L. (2000). Quality, satisfaction and behavioral intentions. *Annals of Tourism Research, 27*(3), 785-804.

Baloglu, S., and Mangaloglu, M. (2001). Tourism destination images of Turkey, Egypt, Greece, and Italy as perceived by US-based tour operators and travel agents. *Tourism Management, 22*(1), 1-9.

Bigne, J. E., Sanchez, M. I., and Sanchez, J. (2001). Tourism image, evaluation variables and after purchase behaviour: inter-relationship. *Tourism Management, 22*(6), 607-616.

Callan, R. J., and Hoyes, M. K. (2000). A preliminary assessment of the function and conference service product at a UK stately home. *Tourism Management, 21*(6), 571-581.

Carmichael, B. A. (2000). A matrix model for resident attitudes and behaviours in a rapidly changing tourist area. *Tourism Management, 21*(6), 601-611.

Chaudhary, M. (2000). India's image as a tourist destination -- a perspective of foreign tourists. *Tourism Management, 21*(3), 293-297.

Chen, J. S. (2001). A case study of Korean outbound travelers' destination images by using correspondence analysis. *Tourism Management, 22*(4), 345-350.

Chen, J. S., and Hsu, C. H. C. (2001). Developing and validating a riverboat gaming impact scale. *Annals of Tourism Research, 28*(2), 459-476.

Chu, R. K. S., and Choi, T. (2000). An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: a comparison of business and leisure travellers. *Tourism Management, 21*(4), 363-377.

de Albuquerque, K., and McElroy, J. L. (2001). Tourist harassment: Barbados Survey Results. *Annals of Tourism Research, 28*(2), 477-492.

Dharmaratne, G. S., Yee Sang, F., and Walling, L. J. (2000). Tourism potentials for financing protected areas. *Annals of Tourism Research, 27*(3), 590-610.

Fairweather, J. R., and R. Swaffield, S. (2001). Visitor Experiences of Kaikoura, New Zealand: an interpretative study using photographs of landscapes and Q method. *Tourism Management, 22*(3), 219-228.

Fleischer, A., & Pizam, A. (2001). Tourism constraint among Israeli seniors. *Annals of Tourism Research, 29*(1), 106-123.

Gilbert, D., and Hudson, S. (2000). Tourism demand constraints: A skiing participation. *Annals of Tourism Research, 27*(4), 906-925.

Gursoy, D., and Chen, J. S. (2000). Competitive analysis of cross cultural information search behavior. *Tourism Management, 21*(6), 583-590.

Herbert, D. (2001). Literary places, tourism and the heritage experience. *Annals of Tourism Research, 28*(2), 312-333.

Heung, V. C. S., Qu, H., and Chu, R. (2001). The relationship between vacation factors and socio-demographic and travelling characteristics: the case of Japanese leisure travellers. *Tourism Management, 22*(3), 259-269.

Jacobsen, J. K. S. (2000). Anti-tourist attitudes: Mediterranean charter tourism. *Annals of Tourism Research, 27*(2), 284-300.

Kim, S., and Littrell, M. A. (2001). Souvenir buying intentions for self versus others. *Annals of Tourism Research, 28*(3), 638-657.

Kozak, M. (2001). Comparative assessment of tourist satisfaction with destinations across two nationalities. *Tourism Management, 22*(4), 391-401.

Kozak, M. (2001). Repeaters' behavior at two distinct destinations. *Annals of Tourism Research, 28*(3), 784-807.

Lee, C.-K. (2000). A comparative study of Caucasian and Asian visitors to a Cultural Expo in an Asian setting. *Tourism Management, 21*(2), 169-176.

Lee, H.-C. (2001). Determinants of recreational boater expenditures on trips. *Tourism Management, 22*(6), 659-667.

Litvin, S. W., and Ng Sok Ling, S. (2001). The destination attribute management model: an empirical application to Bintan, Indonesia. *Tourism Management, 22*(5), 481-492.

Master, H., and Prideaux, B. (2000). Culture and vacation satisfaction: a study of Taiwanese tourists in South East Queensland. *Tourism Management, 21*(5), 445-449.

McGehee, N. G. (2001). Alternative tourism and social movements. *Annals of Tourism Research, 29*(1), 124-143.

Mundet, L., and Ribera, L. (2001). Characteristics of divers at a Spanish resort. *Tourism Management, 22*(5), 501-510.

Murphy, P., Pritchard, M. P., and Smith, B. (2000). The destination product and its impact on traveller perceptions. *Tourism Management, 21*(1), 43-52.

Mykletun, R. J., Crotts, J. C., and Mykletun, A. (2001). Positioning an island destination in the peripheral area of the Baltics: a flexible approach to market segmentation. *Tourism Management, 22*(5), 493-500.

Oh, H., and Hsu, C. H. C. (2001). Volitional degrees of gambling behaviors. *Annals of Tourism Research, 28*(3), 618-637.

Orams, M. B. (2000). Tourists getting close to whales, is it what whale-watching is all about? *Tourism Management, 21*(6), 561-569.

Perez, E. A., & Sampol, C. J. (2000). Tourist expenditure for mass tourism markets. *Annals of Tourism Research, 27*(3), 624-637.

Pizam, A., Uriely, N., and Reichel, A. (2000). The intensity of tourist-host social relationship and its effects on satisfaction and change of attitudes: the case of working tourists in Israel. *Tourism Management, 21*(4), 395-406.

Reichel, A., Lowengart, O., and Milman, A. (2000). Rural tourism in Israel: service quality and orientation. *Tourism Management, 21*(5), 451-459.

Tierney, P. T., Dahl, R., and Chavez, D. (2001). Cultural diversity in use of undeveloped natural areas by Los Angeles county residents. *Tourism Management, 22*(3), 271-277.

Uriely, N., and Reichel, A. (2000). Working tourists and their attitudes to hosts. *Annals of Tourism Research, 27*(2), 267-283.

Uysal, M., Chen, J. S., & Williams, D. R. (2000). Increasing state market share through a regional positioning. *Tourism Management, 21*, 89-96.

Vitterso, J., Vorkinn, M., Vistad, O. I., and Vaagland, J. (2000). Tourist experiences and attractions. *Annals of Tourism Research, 27*(2), 432-450.

Waitt, G. (2000). Consuming heritage: Perceived historical authenticity. *Annals of Tourism Research, 27*(4), 835-862.

Walpole, M. J., and Goodwin, H. J. (2000). Local economic impacts of dragon tourism in Indonesia. *Annals of Tourism Research, 27*(3), 559-576.

Table 1: Constructs studied

		Frequency	Percent
satisfaction		15	35
perceptions		14	33
attitudes		10	23
intentions		4	9
behavior		5	12
performance		3	7
other		23	55
	expenditure	3	7
	travel motivation	3	7
	expectations	2	5
	importance	2	5
	intention to return	2	5
	knowledge	2	5
	authenticity	1	2
	constraints	1	2
	familiarity	1	2
	harassment	1	2
	liking pictures	1	2
	social movements	1	2
	social norm	1	2
	visitation barriers	1	2
	willingness to pay	1	2

Table 2: Answer formats chosen for the constructs under study

	categorical	binary	ordinal	metric	not stated	% ordinal
Satisfaction			14		1	93%
Performance			3			100%
Attitudes		1	8		1	80%
Perceptions		1	13			93%
Intentions			4			100%
Behavior	1	2	1		1	20%
Other constructs	1	1	16	4	1	70%
Main construct studied	*1*	*3*	*33*	*4*	*2*	*77%*

Table 3: Justifications of answer formats chosen

	Frequency	Percent
none	27	63
no justification needed	6	14
prior work using similar scale cited	4	9
preexisting scale used	3	7
someone who justified use of such a scale cited	2	5
reasons stated	1	2

Table 4: Data analytic methods applied

	binary	ordinal	metric	not stated	total
Used frequency-based descriptives	3	15	2	2	22
Used mean-based descriptives	0	16	2	0	18
Used factor analysis	0	14	0	0	14
Used t-tests	0	7	1	0	8
Used analysis of variance	0	8	0	0	8
Used chi-squared test	1	6	0	0	7
Used regression analysis	0	6	1	0	7
Used structural equation modelling	0	6	0	0	6

Table 5: Cross-cultural issues

	High danger of misinterpretation	Low danger of misinterpretation	*Total*
cross-cultural response style investigation not needed	na	na	*18*
cross-cultural response style investigation needed	21	2	*23*