University of Wollongong
Research Online

Faculty of Commerce - Papers (Archive)

Faculty of Business

2008

Cross-cultural comparisons of tourist satisfaction: assessing analytical robustness

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

Bettina Grun
Vienna University of Technology, bettina@uow.edu.au

Huong Le
University of Wollongong

Publication Details

This book chapter was originally published as Dolnicar, S, Grün, B & Le, H, Cross-cultural comparisons of tourist satisfaction: assessing analytical robustness, in: Yuksel, A (ed.), Tourist Satisfaction and Complaining Behaviour: Measurement & Management Issues in the Tourism and Hospitality Industry, New York: Nova Publishing Sciences, 2008, 137-150.

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

Cross-cultural comparisons of tourist satisfaction: assessing analytical robustness

Abstract
Response styles can distort survey findings. Culture-specific response styles (CSRS) are particularly problematic to cross-cultural and empirical tourism researchers using multi-cultural samples because the resulting data contamination can lead to inaccurate conclusions about the research question under study. This is particularly the case when constructs such as satisfaction are measured, which are difficult to operationalise. Nevertheless, possible culture-specific response style effects are typically ignored, thus jeopardizing the validity of reported findings. This chapter raises awareness of the problem, illustrates the problem empirically and presents a method that enables researchers to assess the robustness of empirical findings on cross-cultural differences in satisfaction to CSRS. This approach avoids the disadvantages of ignoring the problem and interpreting spurious results or choosing one single correction technique that potentially introduces new kinds of data contamination.

Disciplines
Business | Social and Behavioral Sciences

Publication Details
This book chapter was originally published as Dolnicar, S, Grün, B & Le, H, Cross-cultural comparisons of tourist satisfaction: assessing analytical robustness, in: Yuksel, A (ed.), Tourist Satisfaction and Complaining Behaviour: Measurement & Management Issues in the Tourism and Hospitality Industry, New York: Nova Publishing Sciences, 2008, 137-150.

This book chapter is available at Research Online: http://ro.uow.edu.au/commpapers/521

Cross-Cultural Comparisons of Tourist Satisfaction: Assessing Analytical Robustness

Sara Dolnicar

School of Management & Marketing
marketing research innovation centre (mric)
University of Wollongong
Wollongong, NSW 2522, Australia
Telephone: (61 2) 4221 3862, Fax: (61 2) 4221 4154
sara_dolnicar@uow.edu.au

Bettina Grün

Department of Statistics and Probability Theory
Vienna University of Technology
Wiedner Hauptstraße 8-10/1071, A-1040 Vienna, Austria
Telephone: (43 1) 58801 10716, Fax: (43 1) 58801 10798
bettina.gruen@ci.tuwien.ac.at

Huong Le

School of Management & Marketing
University of Wollongong
Wollongong, NSW 2522, Australia
Telephone: (61 2) 4221 4772, Fax: (61 2) 4221 4154
huong_le@uow.edu.au

Cross-Cultural Comparisons of Tourist Satisfaction: Assessing Analytical Robustness

Abstract

Response styles can distort survey findings. Culture-specific response styles (CSRS) are particularly problematic to cross-cultural and empirical tourism researchers using multi-cultural samples because the resulting data contamination can lead to inaccurate conclusions about the research question under study. This is particularly the case when constructs such as satisfaction are measured, which are difficult to operationalise. Nevertheless, possible culture-specific response style effects are typically ignored, thus jeopardizing the validity of reported findings. This chapter raises awareness of the problem, illustrates the problem empirically and presents a method that enables researchers to assess the robustness of empirical findings on cross-cultural differences in satisfaction to CSRS. This approach avoids the disadvantages of ignoring the problem and interpreting spurious results or choosing one single correction technique that potentially introduces new kinds of data contamination.

1 Introduction

The construct of tourist satisfaction has been studied extensively in the past. Yet, very little research has been undertaken to assess how tourist satisfaction can most validly be measured or which measures may be prone to bias. One of the few studies investigating such effects was published by Sirakaya, Petrick and Choi (2004). The authors find that the mood of respondents affects satisfaction rating. Yet, mood can be claimed to be a variable which is likely to be randomly distributed across the sample. As such the bias of mood is likely to even out across all respondents in the sample. Factors of real concern, however, are those that are systematically associated with certain respondents, such as age, gender and education level, and that are also of interest in the analysis. One such variable has been repeatedly identified as causing

systematic bias in survey responses: the cultural background of respondents. Pizam and Ellis (1999) discuss these "global issues" in consumer satisfaction measurement in detail. They identify a large number of potential biases that can distort satisfaction data collected from respondents from different cultural backgrounds. The particular aspect that will be discussed in this book chapter is referred to as "scalar equivalence" with the key question Pizam and Ellis recommend satisfaction researchers should ask being "Do corporate chosen scales function similarly in different cultures?" (p. 336).

The aim of this chapter is to raise awareness for the problem of scalar equivalence in satisfaction measurement among tourism and hospitality researchers, to empirically demonstrate the problem and to illustrate a simple method that can help researchers assess how robust their findings regarding the identified cross-cultural differences are.

2 Response styles

Throughout this chapter the term *response bias* will be understood to be "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)". Furthermore, the term *response style* will refer to response bias that "an individual displays […] consistently across time and situations" (Paulhus, 1991, p.17). The two main forms of response styles are Extreme Response Style (ERS) and Acquiescence Response Style (ARS). Respondents with an ERS tend to use the endpoints of an answer scale. Respondents with an ARS tend to give a positive answer.

Substantial empirical evidence exists for the fact that the cultural background of respondents heavily affects the way in which they use answer formats in questionnaires. Zax and Takahashi (1967) conducted one of the earliest empirical studies on cross-cultural response styles concluding that Japanese female students exhibit ERS to a higher extent than their American counterparts. Chun, Campbell and Yoo (1974) conclude that American respondents demonstrate higher ERS scores than Korean participants in surveys, whereas Marshall and Lee (1998) find that in a

comparison of seven Asian and Western countries the Asian respondents have a higher level of ERS. Differences in responses styles have also been empirically demonstrated to exist between respondents from different European countries (van Herk, Poortinga, & Verhallen, 2004; Welkenhuysen-Gybels, Billiet & Cambre, 2003) and between Hispanic and Non-Hispanic respondents (Hui & Triandis, 1989; Marin, Gamba, & Marin 1992) generally concluding that Hispanic respondents are to a higher extent susceptible to ERS.

It should be noted at this point that most empirical studies demonstrating response styles have used multi-category answer formats, such as five or seven point scales which currently dominate empirical social science research. As early as in 1950 Cronbach (1950) – aware of the serious problem of response styles for the validity of survey findings – recommended to use binary scales with only two answer options to avoid the contamination of data with not content related systematic error. Clarke III (2000, 2001) provides some empirical support for Cronbach's early recommendation. He finds that using a higher number of scale options is more susceptible to culturally determined response styles.

Although – to the authors' knowledge – no empirical work has been done to better understand why respondents from different cultural backgrounds use answer formats in a different manner, a few of the authors of the above cited empirical studies propose some explanations: Hui and Triandis (1989) propose that the difference between cultures lies in how they match the continuous construct that is being examined by the questions with the limited number of answer categories available in a questionnaire. The argument made by Stening and Everett (1984) is based on difference in value systems. In Asia modesty is an important trait. Using extreme response options is not modest. This may be a reason that Asian respondents are known to tend to use the middle answer options. Contrarily, Hispanic respondents believe that questions must be answered honestly. Honesty is expressed by taking strong positions and using the endpoints on an answer scale more frequently. More generally, Pizam and Ellis (1999, p. 335) state that "Differing languages, levels of

literacy, interpretations of constructs and cultural behaviour must all be taken into account when creating a foreign customer satisfaction survey."

The best way of addressing the problem clearly is to collect data in a way that is not susceptible to capturing response styles. This leads back to Cronbach's recommendation of considering to use binary answer formats. In addition new answer formats such as best-worst scaling can be used for certain kinds of questions. Lee, Soutar, Louviere and Daly (2006) used best-worst scaling and could not detect any cross-cultural response styles in their data. This is, however, not always possible. If data sets have already been collected or if the researcher only has limited influence on the questionnaire development it may be necessary to work with data that is likely to be contaminated by response styles. A number of authors have made recommendations how to detect and correct for response styles before conducting the analysis (Fischer, 2004; Byrne & Campbell, 1999; Cheung & Rensvold, 2000; Greenleaf, 1992a and b; Van de Vijver & Poortinga, 2002; Welkenhuysen-Gybels, Billiet & Cambre, 2003). The proposed methods range from simple counting procedures to modelling approaches to extract ERS and ARS. All of the correction approaches proposed, however, have one major disadvantage: they assume that they have detected the true nature of the response style which they subsequently eliminate. Any data transformation is endangered by being either incorrect or introducing new biases into the data. The method proposed by Dolnicar and Grün (2007) is illustrated in this chapter avoids this problem. It is a diagnostic tool that informs researchers about the robustness of their results and therefore protects them from drawing wrong conclusions without manipulating the original data set.

3 Response styles in tourism satisfaction research

To assess the extent to which satisfaction research in the field of tourism is affected by the problem of response styles, a descriptive bibliography study [1] was conducted.

The following journals were used as sources for publications on tourism satisfaction: Journal of Travel Research, Annals of Tourism Research, Tourism Management, International Journal of Hospitality Management, Cornell Hotel and Restaurant Administration Quarterly, and the Journal of Tourism Studies. These journals were selected because they are ranked among the top 10 journals in tourism research according to the tourism journal ranking published by the Journal of Travel Research in 2004 and because they are readily available through online databases.

Articles to be included in the review were selected by searching for the keyword "satisfaction". Only original articles based on empirical satisfaction data and published between 2000 and 2007 were included. This selection algorithm led to a total of 45 articles used for the review. The full list of references is provided in the Appendix. The distribution of papers across publication outlets is provided in Table 1. Table 1 shows that Tourism Management published the largest number of empirical satisfaction studies, more than half of all studies undertaken in the listed journals since 2000. The Journal of Travel Research published the second largest number of satisfaction studies, followed by Annals of Tourism Research. Counting the total number of articles published in the top three tourism journals and the number of articles which investigate satisfaction, it becomes evident that a substantial amount of satisfaction research is published in tourism. In Tourism Management, 11% of all articles published from 2000 study satisfaction, the respective proportions for the Journal of Travel Research and Annals of Tourism Research are 8% and 6%.

---

[1] Bibliographic study (also called bibliographical study) is a systematic description and history of printed material (Center for Bibliographical Studies and Research, 2006).

The review of articles was undertaken by coding each article with respect to a set of predefined variables. Twenty six such variables were used. Variables were divided into three groups: (1) aims of the research and general aspects of articles such as authors; years of publication; names of the journals; (2) aspects of methodology such as sample size determination; if the sample involved people from one country or from different countries, if the authors correct for cross-cultural response styles, number of attributes, data format, number of answer options, statistical analysis; (3) measurement aspects such as how and when levels of tourists' satisfaction were measured, if importance of attributes was measured, measurement of behavioural intentions as consequences of satisfaction; and (4) the main results of the research and managerial recommendations or managerial notes.

Table 1 – Distribution of empirical satisfaction studies across journals

Journal of publication	**Total number**	**%**
Tourism Management	24.0	53.3
Journal of Travel Research	10.0	22.2
Annals of Tourism Research	7.0	15.6
International Journal of Hospitality Management	2.0	4.4
Cornell Hotel and Restaurant Administration Quarterly	1.0	2.2
Journal of Tourism Studies	1.0	2.2
Total	45.0	100.0

Table 2 contains the results of the review for the main variables of interest. As can be seen, more than half of all satisfaction studies conducted in tourism research use data sets that contain respondents from more than one country. This is not surprising because frequently guest surveys are used to study satisfaction. Yet, it highlights the importance of addressing the issue of potential data contamination by response styles. The risk of data contamination by response styles has to be assessed as high given that 93 percent of all studies use multi-category scales to measure satisfaction: nearly half of the studies chose five-point scales, almost a third of studies use seven-point scales. None of the studies that include respondents from different countries of origin have corrected for cross-cultural response styles. In fact, none of

them even mention that cross-cultural response styles could potentially bias the results.

Table 2 – Characteristics of satisfaction studies in tourism research

	Frequency	**Percent**
Respondents		
Only from 1 country	14	31.1
From more than 1 country	24	53.3
Not stated	7	15.6
Total	*45*	*100.0*
Correction for cross-cultural response styles		
No	45	100.0
Yes	0	0.0
Total	*45*	*100.0*
Data format		
Multi-category	42	93.3
Not stated	1	2.2
Not applicable (qualitative studies)	2	4.4
Total	*45*	*100.0*
Number of answer options		
Not specified	2	4.4
4 point scale	1	2.2
5 point scale	19	42.2
6 point scale	1	2.2
7 point scale	13	28.9
9 point scale	1	2.2
10 point scale	4	8.9
Others / combination of formats	2	4.4
Not applicable (qualitative studies)	2	4.4
Total	*45*	*100.0*

Given the concerning statistics presented in Table 2, a more detailed analysis of those articles that used respondents from various countries of origin was undertaken, leading to the conclusion that 7 out of 45 of the reviewed studies (18%) actually examined cross-cultural differences in satisfaction levels. Nield, Kozak &

LeGrys (2000), in their study on satisfaction of tourists with food use five point scales to measure satisfaction and compare Western European with Eastern Europeans. Chaudhary (2000) compares satisfaction ratings on five point scales for British, German and Dutch tourists. Results are insignificant, the authors blame small sample sizes, and response styles problems are not mentioned. Kozak (2001) compares the satisfaction statements of British and German tourists using a seven point scale. Joppe, Martin and Waalen (2001) use four and five point scales and compare Canadian, US and overseas tourist satisfaction levels. Wong & Law (2003) compare satisfaction levels across countries of origin using a five point scale to measure satisfaction. Yu & Goulden (2006) test differences in satisfaction of European, US, Japanese and other Asia Pacific Countries using satisfaction statements measured on a 5 point multicategory scale. Hui, Wan and Ho (in press) compare satisfaction levels for respondents from different regions of the world. The basis for the analysis are responses provided on a seven point scale.

With the exception of Chaudhary (2000) all studies report significant differences across countries. In many cases these differences are not only significant, they are obviously highly systematic with respondents from certain cultural backgrounds producing higher satisfaction scores consistently over a large number of attributes for which satisfaction was measured. Not a single one of these studies mentions the potential danger of cross-cultural responses styles.

The results from this bibliographic study demonstrate very clearly that tourism researchers are in need of a tool that will enable them to discriminate between response style artefacts and true cross-cultural differences. We illustrate the problem of cross-cultural response style contamination and a simple method to assess the danger of data contamination in the following section.

4 An empirical illustration

The data set used for the empirical illustration is from the most recent wave (1999-2002) of the World Values Study (Inglehart, Basanez, Diez-Medrano, Halman

& Luijkx, 2004), a data set collected by a network of social scientists since 1980. Random sampling techniques are used in all countries and only respondents of the age of 18 and above are included.

The analysis is restricted only to a subset containing the respondents of three different countries (n=3771): United States of America (1200 respondents, 32% of the sample), Spain (1209 respondents, 32%) and Japan (1362 respondents, 36%). These three countries are chosen because cross-cultural analyses of response styles have often been made between Americans and Hispanic as well as Asian respondents and have shown significant differences in response styles. Consequently, it is reasonable to assume that respondents from these countries will differ in the way they respond to multi-category survey questions.

Sixty-seven questions from the World Values Survey form the basis of the analysis, 47 of which respondents answered by using a four point scale. Respondents answered the remaining 20 questions on a ten point scale. The core variables for analysis are four questions which investigate the satisfaction of respondents. More specifically, the following aspects of satisfaction: (1) satisfaction with life (four point scale), (2) satisfaction with the financial situation of the household (four point scale), (3) satisfaction with democracy developing in their country of residence (ten point scale) and (4) satisfaction with the people in the national office (ten point scale).

The answer options for the first questions were “Very satisfied”, “Rather satisfied”, “Not very satisfied”, and “Not at all satisfied”; for the second question “Very satisfied”, “Fairly satisfied”, “Fairly dissatisfied” and “Very dissatisfied”. Both ten point scale questions required respondents to use a numerical scale with the endpoints anchored verbally as “Dissatisfied” and “Satisfied”.

The satisfaction questions in the World Values Data represent very well the nature of questions typically asked when satisfaction is measured in a tourism context. The typical approach to testing whether respondents from different countries of origin

have different satisfaction levels is to conduct ANOVAs and establish whether the mean values differ. In this case the ANOVA for each of the four satisfaction question indicates a significant difference (all p-values < 0.001). In the next step pair-wise comparisons are made using Tukey's honest significant different (HSD) method to correct for multiple testing in order to assess which countries differ significantly. Each one of these pair-wise tests has three possible outcomes: (1) respondents from country A are more satisfied (A>B), (2) there is no difference in the satisfaction of respondents from countries A and B (A=B), and (3) respondents from country B are more satisfied (A<B). Ignoring the problem of cross-cultural response styles these results (see Table 3 providing the mean differences and p-values for all pair-wise comparisons of the three countries along all four satisfaction variables) would be interpreted as follows: Japanese respondents are for the least satisfied (across all items), and Americans are the most satisfied (across all items except the questions regarding democracy). However, the possibility that these differences are systematic and that we may in fact be interpreting response styles rather than actual content has not been taken into account by this analysis, consequently putting the results at risk of being invalid.

Table 3: Analysis of the Raw Data

Answer Scale	Question	Spain vs. Japan (p-value)	USA vs. Japan (p-value)	USA vs. Spain (p-value)
Ten point	Life	0.109 A>B (< 0.001)	0.261 A>B (< 0.001)	0.153 A>B (< 0.001)
	Financial situation of household	0.020 A=B (0.56)	0.075 A>B (< 0.001)	0.055 A>B (0.02)
Four point	Democracy developing in country	0.253 A>B (< 0.001)	0.206 A>B (<0.001)	-0.047 A<B (0.03)
	People in the national office	0.370 A>B (< 0.001)	0.541 A>B (< 0.001)	0.171 A>B (< 0.001)

In order to address this problem the presence of cross-cultural response styles is investigated. Individual means and standard deviations are the recommended measures for assessing ARS and ERS respectively. For instance, if a respondent has high agreement levels for all satisfaction questions, states to have engaged in many vacation activities and states that most travel motivations apply to him or her (including resting, relaxing and doing nothing), the validity of his or her responses is in question, as an observed general tendency of using only the positive range of the scale over several different constructs is more likely to be a sign of ARS than of actual content information. As a consequence his or her mean value over all questions will be rather high and hence reflect the degree of susceptibility to ARS of the respondent. Similar, the observed individual standard deviation over several questions from unrelated constructs is used as a measure for susceptibility to ERS of the respondent.

For the present illustration individual mean values and standard deviations are determined separately for each answer format because previous research demonstrated that the susceptibility of answer formats to culture-specific response styles is associated with the number of answer categories (Hui & Triandis, 1989; Clarke III, 2000, 2001). In order to determine if the individual means and standard deviations are valid measures for response styles the interdependence between the different questions in the questionnaire is analyzed. Low correlations between the answers suggest that they are unrelated and systematic differences in use of the scale between the respondents are likely to be due to response styles. Given that the correlations have a mean of 0.07 (standard deviation 0.13) for the four point scale and

a mean of 0.06 (standard deviation 0.19) for the ten point scale it can be assumed that the individual means and standard deviations can be used as measures for ARS and ERS.

In order to assess cross-cultural differences in response styles ANOVAs are performed which show that the countries differ significantly in the individual means and standard deviations (four point scale: F-value=124, df_1=2, df_2=3768, p-value < 0.001 (means), F-value=106, df_1=2, df_2=3768, p-value < 0.001 (standard deviations); ten point scale: F-value=84, df_1=2, df_2=3763, p-value < 0.001 (means); F-value=61.1, df_1=2, df_2=3763, p-value < 0.001 (standard deviations)). A pair-wise comparison using Tukey's HSD method at a significance level of 95% indicates that the Americans have the highest means and the Japanese the lowest means and standard deviations, while the difference in standard deviations are not significant between Americans and Spanish for the four point scale. For the ten point scale the Japanese have again the lowest means and standard deviations, while the differences in means are not significant between Americans and Spanish. However, the Americans have higher standard deviations than the Spanish. These results suggest that analyzing uncorrected raw data might be distorted by the presence of culture-specific response styles.

The seemingly logical consequence from the above results is that the raw data needs to be corrected; that the scores for each of the three cultural groups have to be somehow modified to reduce the amount of bias. Unfortunately this is a dangerous approach. By correcting the raw data additional or different bias could be introduced to the data. The approach we are illustrating in this book chapter therefore does not take a correction approach. Instead we present a simple way to assess how reliable each of the differences are that we originally found between respondents from the three countries. We refer to this as a robustness comparison. A detailed explanation of the procedure is provided in Dolnicar and Grün (2007). The underlying idea is that we apply a number of alternative, theoretically suitable corrections to the data, re-compute the original test to compare the three countries and then assess whether the results from the different correction methods as well as the raw data lead to the same

or different results. If all of them lead to the same result (either that the countries differ or that they do not in their satisfaction) we can safely assume that this is the correct result, despite the response style contamination. If, however, there is no agreement on the results, findings with respect to such a variable have to be reported with great care, as it cannot be firmly established if a satisfaction difference or a response style difference is captured.

For the robustness comparison of the World Values Data we use seven different correction methods: the raw data, the data corrected for ARS using individual means as well as using country-specific means, the data corrected for ERS using individual standard deviations as well as using country-specific standard deviations and the data corrected for both ARS and ERS using either the individual measures as well as the country-specific ones.

The ANOVA indicates that country-specific differences indeed do exist for each of the questions (all p-values < 0.001 for each corrected data set and question). This preliminary result is very encouraging, as it confirms that the identified differences are not merely based on response styles. However, pair wise comparisons are needed to be able to draw final conclusions about possible cross-cultural differences. For this purpose Tukey's HSD method was used (significance level of 95%). Because the test is computed seven times (once for each of the corrected data sets and once for the raw data) each of the three cases can occur between 0 and 7 times. The higher the agreement across the seven computations are the more robust the finding. Optimally the resulting values will mainly be 0s and 7s. In the worst case most of them will be 3s and 4s, indicating high levels of correction dependence of results.

Table 4 contains the results of these pair-wise tests for the World Values Data. The respective country pair is stated in the column heading. Each row contains the frequency of the three outcomes as outlined above (A>B, A=B, A<B) for each satisfaction item under study. The robust test results are highlighted in a light grey

shade. As can be seen five of six comparisons on the four point scale are highly robust, but only two of six on the ten point scale.

Table 4: Robustness of Cross-Cultural Findings

Answer Scale	Question	Spain vs. Japan	USA vs. Japan	USA vs. Spain
Ten point	Life	A<B 0 A=B 4 A>B 3	A<B 0 A=B 0 A>B 7	A<B 0 A=B 0 A>B 7
	Financial situation of household	A<B 4 A=B 3 A>B 0	A<B 0 A=B 4 A>B 3	A<B 0 A=B 3 A>B 4
Four point	Democracy developing in country	A<B 0 A=B 0 A>B 7	A<B 0 A=B 0 A>B 7	A<B 6 A=B 1 A>B 0
	People in the national office	A<B 0 A=B 0 A>B 7	A<B 0 A=B 0 A>B 7	A<B 0 A=B 0 A>B 7

None of the pair wise comparisons indicates unambiguously insignificant differences between two countries. The comparisons indicate that the Japanese are the least satisfied with respect to the democracy developing in their country. No safe conclusion for this question can be drawn for the comparison of Americans and Spanish respondents, as the data set corrected for individual standard deviations indicates no significant differences between these two countries (p-value=0.45). With respect to satisfaction with the people in the national office the Japanese are again the least satisfied and the Americans are the most satisfied. With respect to satisfaction with their life the Americans are the most satisfied and with respect to satisfaction with the financial situation of the household no safe conclusions can be drawn for any of the comparisons. A majority vote of the corrected data sets would indicate lower levels for Spanish than Japanese and Americans while insignificant differences are suggested between the Americans and the Japanese. The majority vote would therefore agree with the raw data analysis only for one out of four comparisons which are assessed as not robust for the ten point scale. However, these conclusions would also not seem to be very reliable as the majority vote is always only based on 4 out of 7 corrected data sets.

The analysis of satisfaction questions from the World Values Data illustrates that response styles can have a major distorting effect on cross-cultural studies. In the worst case response styles can lead to wrong conclusions. It is consequently very important for researchers who are interested in comparing satisfaction ratings from

respondents from different countries of origin to assess the degree to which their results are based on differences in satisfaction (actual content) or differences in using answer formats (response styles).

5 Conclusions

Satisfaction research is very popular among tourism researchers. Satisfaction is assumed to play a central role in tourists' intentions to revisit a destination and to lead to positive word of mouth. The majority of satisfaction studies use multi-category answer formats to measure satisfaction, either directly or through the measurement of both expectations / importance and performance independently. A large proportion of satisfaction studies is based on data sets which include respondents from different cultural backgrounds who are known to use multi-category response scales in systematically different ways. Such systematic differences can affect the validity of conclusions drawn from empirical satisfaction research, particularly if respondents from different cultural backgrounds are directly compared. In the worst case – if researchers are comparing countries with very strong response styles – the statistically significant differences in satisfaction as determined by an analysis of variance or t-test may be entirely due to differences in response styles. This would mean that tourists from different countries do not at all differ in their satisfaction. In addition true differences in satisfaction can also be masked by response styles and hence might not be detected.

Because of the danger of interpreting methodological artefacts it is particularly important in the context of empirical tourism research to assess the extent of the potential contamination of data with response styles. One way of doing this was presented in this chapter: first the raw data is corrected for various possible response styles. The derived data set and the raw data set are used independently to undertake significance testing. For each variable, the test results of all (raw and corrected) data sets are compared. The higher the level of agreement between those computations, the more reliable the finding that countries do or do not significantly differ with respect to that particular aspect of satisfaction.

Tourism research may also want to consider alternative answer formats, such as best-worst scaling or binary answer formats in cases where these answer formats are viable ways of collecting the required data. Best-worst scales and binary scales are less susceptible to capturing response styles than the typically used multi-category answer formats.

Within the group of multi-category answer formats and their susceptibility to response styles, future empirical studies are needed. Particularly to assess whether lower number of answer options are generally more robust to culture-specific response styles than ten point scales (as suggested by our empirical analysis where four point scales appeared to be less in danger than ten point scales). In addition the effect of labeling of answer formats on cross-cultural response style susceptibility needs to be assessed empirically.

6 Acknowledgements

This research was supported by the Australian Research Council (through grants DP0557257 and LX0559628) and the Austrian Academy of Sciences (ÖAW) through a DOC-FFORTE scholarship for Bettina Grün.

References

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.

Center for Bibliographical Studies and Research (2006). *Website of the Center for Bibliographical Studies and Research (CBSR)*. Retrieved 2 January, 2007 from http://www.cbsr.ucr.edu/

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modelling. *Journal of Cross-Cultural Psychology, 31*(2), 187-212.

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

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.

Clarke III, I. (2000). Extreme response style in cross cultural research: An empirical investigation. *Journal of Social Behaviour and Personality, 15*(1), 137-152.

Clarke III, I. (2001). Extreme response style in cross-cultural research. *International Marketing Review, 18*(3), 301-324.

Cronbach, L. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement, 10* (1), 3-31.

Dolnicar, S. & Grün, B. (2007). Cross-cultural comparisons: Assessing analytical robustness. *International Journal of Culture, Tourism and Hospitality Research*.

Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology, 35*(3), 263-282.

Greenleaf, E. A. (1992a). Improving rating-scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research, 29*(2), 176-188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly, 56*(3), 328-351.

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.

Hui, T. K., Wan, D., & Ho, A. (in press). Tourists' satisfaction, recommendation and revisiting Singapore. *Tourism Management, In Press, Corrected Proof.*

Inglehart R., Basanez M., Diez-Medrano J., Halman L., Luijkx R. (2004). *Human Beliefs and Values: A Cross-Cultural Sourcebook Based on the 1999-2002 Values Surveys*. Mexico: Siglo XXI Editores.

Joppe, M., Martin, D. W., & Waalen, J. (2001). Toronto's image as a destination: A comparative importance-satisfaction analysis by origin of visitor. *Journal of Travel Research, 39*(3), 252-260.

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

Lee, J. A., Soutar, G. N., Louviere, J., & Daly, T. M. (2006). *An Examination of the Relationship between Values and Holiday Benefits Across Cultures Using Ratings Scales and Best-Worst Scaling*. ANZMAC CD Proceedings.

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. *European Advances in Consumer Research*, *3*, 90-95.

Nield, K., Kozak, M., & LeGrys, G. (2000). The role of food service in tourist satisfaction. *International Journal of Hospitality Management, 19*(4), 375-384.

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*, San Diego: Academic Press, 17-59.

Pizam, A., & Ellis, T. (1999). Customer satisfaction and its measurement in hospitality enterprises. *International Journal of Contemporary Hospitality Management, 11*(7), 326-339.

Sirakaya, E., Petrick, J., & Choi, H.-S. (2004). The role of mood on tourism product evaluations. *Annals of Tourism Research, 31*(3), 517-539.

Stening, B. W., & Everett, J. E. (1984). Response styles in a cross-cultural managerial study. *Journal of Social Psychology, 122*(2), 151-156.

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 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.

Wong, J., & Law, R. (2003). Difference in shopping satisfaction levels: A study of tourists in Hong Kong. *Tourism Management, 24*(4), 401-410.

Yu, L., & Goulden, M. (2006). A comparative analysis of international tourists' satisfaction in Mongolia. *Tourism Management, 27*(6), 1331-1342.

Zax, M., & Takahashi, S. (1967), Cultural influences on response style: Comparison of Japanese and American college students. *Journal of Social Psychology, 71*(1), 3-10.

Appendix: Reviewed literature

Akama, J. S., & Kieti, D. M. (2003). Measuring tourist satisfaction with Kenya's wildlife safari: a case study of Tsavo West National Park. *Tourism Management, 24*(1), 73-81.

Alegre, J., & Cladera, M. (2006). Repeat visitation in mature sun and sand holiday destinations. *Journal of Travel Research, 44*(3), 288-297.

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

Bigne, J. E., & Andreu, L. (2004). Emotions in segmentation: An empirical study. *Annals of Tourism Research, 31*(3), 682-696.

Bigne, J. E., Andreu, L., & Gnoth, J. (2005). The theme park experience: An analysis of pleasure, arousal and satisfaction. *Tourism Management, 26*(6), 833-844.

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

Bowen, D. (2001). Antecedents of consumer satisfaction and dis-satisfaction (CS/D) on long-haul inclusive tours -- A reality check on theoretical considerations. *Tourism Management, 22*(1), 49-61.

Bowen, D. (2002). Research through participant observation in tourism: A creative solution to the measurement of consumer satisfaction/dissatisfaction (CS/D) among tourists. *Journal of Travel Research, 41*(1), 4-14.

Chang, J., Yang, B.-T., & Yu, C.-G. (2006). The moderating effect of salespersons' selling behaviour on shopping motivation and satisfaction: Taiwan tourists in China. *Tourism Management, 27*(5), 934-942.

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

Chen, C.-F., & Tsai, D. (in press). How destination image and evaluative factors affect behavioural intentions? *Tourism Management, In Press, Corrected Proof.*

Deng, W. (in press). Using a revised importance-performance analysis approach: The case of Taiwanese hot springs tourism. *Tourism Management, In Press, Corrected Proof.*

Duman, T., & Mattila, A. S. (2005). The role of affective factors on perceived cruise vacation value. *Tourism Management, 26*(3), 311-323.

Ekinci, Y., Prokopaki, P., & Cobanoglu, C. (2003). Service quality in Cretan accommodations: Marketing strategies for the UK holiday market. *International Journal of Hospitality Management, 22*(1), 47-66.

Fuchs, M., & Weiermair, K. (2004). Destination benchmarking: An indicator-system's potential for exploring guest satisfaction. *Journal of Travel Research, 42*(3), 212-225.

Gallarza, M. G., & Gil Saura, I. (2006). Value dimensions, perceived value, satisfaction and loyalty: An investigation of university students' travel behaviour. *Tourism Management, 27*(3), 437-452.

Heung, V. C. S., & Cheng, E. (2000). Assessing tourists' satisfaction with shopping in the Hong Kong special administrative regions of China. *Journal of Travel Research, 38*(4), 396-404.

Hui, T. K., Wan, D., & Ho, A. (in press). Tourists' satisfaction, recommendation and revisiting Singapore. *Tourism Management, In Press, Corrected Proof.*

Jang, S., & Feng, R. (2007). Temporal destination revisit intention: The effects of novelty seeking and satisfaction. *Tourism Management, 28*(2), 580-590.

Joppe, M., Martin, D. W., & Waalen, J. (2001). Toronto's Image as a destination: A comparative importance-satisfaction analysis by origin of visitor. *Journal of Travel Research, 39*(3), 252-260.

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

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

Kozak, M. (2002). Destination benchmarking. *Annals of Tourism Research, 29*(2), 497-519.

Kozak, M., & Rimmington, M. (2000). Tourist satisfaction with Mallorca, Spain, as an off-season holiday destination. *Journal of Travel Research, 38*(3), 260-269.

Lee, C.-K., Yoon, Y.-S., & Lee, S.-K. (2007). Investigating the relationships among perceived value, satisfaction, and recommendations: The case of the Korean DMZ. *Tourism Management, 28*(1), 204-214.

Millan, A., & Esteban, A. (2004). Development of a multiple-item scale for measuring customer satisfaction in travel agencies services. *Tourism Management, 25*(5), 533-546.

Nash, R., Thyne, M., & Davies, S. (2006). An investigation into customer satisfaction levels in the budget accommodation sector in Scotland: A case study of backpacker tourists and the Scottish Youth Hostels Association. *Tourism Management, 27*(3), 525-532.

Nield, K., Kozak, M., & LeGrys, G. (2000). The role of food service in tourist satisfaction. *International Journal of Hospitality Management, 19*(4), 375-384.

Petrick, J. F. (2004). The roles of quality, value, and satisfaction in predicting cruise passengers' behavioural intentions. *Journal of Travel Research, 42*(4), 397-407.

Petrick, J. F., & Backman, S. J. (2002). An examination of the determinants of golf travelers' satisfaction. *Journal of Travel Research, 40*(3), 252-258.

Pizam, A., Uriely, N., & 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.

Reisinger, Y., & Turner, L. W. (2002a). Cultural differences between Asian tourist markets and Australian hosts: Part 1. *Journal of Travel Research, 40*(3), 295-315.

Reisinger, Y., & Turner, L. W. (2002b). Cultural differences between Asian tourist markets and Australian hosts: Part 2. *Journal of Travel Research, 40*(4), 374-384.

Reisinger, Y., & Turner, L. W. (2002c). The determination of shopping satisfaction of Japanese tourists visiting Hawaii and the Gold Coast compared. *Journal of Travel Research, 41*(2), 167-176.

Rodriguez del Bosque, I. A., San Martin, H., & Collado, J. (2006). The role of expectations in the consumer satisfaction formation process: Empirical evidence in the travel agency sector. *Tourism Management, 27*(3), 410-419.

Sirakaya, E., Petrick, J., & Choi, H.-S. (2004). The role of mood on tourism product evaluations. *Annals of Tourism Research, 31*(3), 517-539.

Skogland, I., & Siguaw, J. A. (2004). Are your satisfied customers loyal? *Cornell Hotel and Restaurant Administration Quarterly, 45*(3), 221-234.

Spinks, W., Lawley, M., & Richins, H. (2005). Satisfaction with Sunshine Coast tourist attractions: The influence of individual visitor characteristics. *Journal of Tourism Studies, 16*(1), 12-23.

Tonge, J., & Moore, S. A. (in press). Importance-satisfaction analysis for marine-park hinterlands: A Western Australian case study. *Tourism Management, In Press, Corrected Proof.*

Truong, T.-H., & Foster, D. (2006). Using HOLSAT to evaluate tourist satisfaction at destinations: The case of Australian holidaymakers in Vietnam. *Tourism Management, 27*(5), 842-855.

Um, S., Chon, K., & Ro, Y. (2006). Antecedents of revisit intention. *Annals of Tourism Research, 33*(4), 1141-1158.

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

Wong, J., & Law, R. (2003). Difference in shopping satisfaction levels: A study of tourists in Hong Kong. *Tourism Management, 24*(4), 401-410.

Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. *Tourism Management, 26*(1), 45-56.

Yu, L., & Goulden, M. (2006). A comparative analysis of international tourists' satisfaction in Mongolia. *Tourism Management, 27*(6), 1331-1342.

Yuksel, A., & Yuksel, F. (2007). Shopping risk perceptions: Effects on tourists' emotions, satisfaction and expressed loyalty intentions. *Tourism Management, 28*(3), 703-713.