Cross-cultural differences in survey response patterns

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

This article has been accepted by Emerald and will be published as: Dolnicar, S & Grun, B, Cross cultural differences in survey response patterns, International Marketing Review. The journal homepage is available here.

Cross-cultural differences in survey response patterns

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

Bettina Grün*
Research Fellow
Department of Statistics and Probability Theory
Vienna University of Technology
Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria
Telephone: (43 1) 58801 10716, Fax: (43 1) 58801 10798, bettina.gruen@ci.tuwien.ac.at

*Authors listed in alphabetical order.

Submission date: 17.11.2005

Revision date: 14.7.2006

Acceptance date: 9.8.2006

Cross-cultural differences in survey response patterns

Structured Abstract

Purpose

The existence of variable response styles represents a major threat to the correct interpretation of market research findings. In international marketing this threat is further increased due to samples of respondents from different cultural backgrounds. In this paper we (1) extend the investigation of differences in cross-cultural response styles by studying full response patterns instead of extreme values, (2) quantify the extent of the potential mistake of not accounting for cross-cultural differences in response behaviour, and (3) present a simple way of testing whether or not data sets from various cultural backgrounds can be used without correcting for cross-cultural response styles.

Methodology

Two independent data sets are used. Extreme response style (ERS) scores are compared by testing for equality of proportions. Respondents' answer patterns are partitioned using the *k*-means algorithm (Hartigan & Wong, 1979), the resulting differences between cultures tested using a Fisher's exact test for count data. The extent of inter-cultural difference in responses is assessed using ANOVA.

Findings

Asian and Australian respondents differ significantly in ERS and full response patterns. Differences in cross-cultural response patterns account for up to six percent of the variance in the data, thus representing a significant potential source for misinterpretation in cross-cultural studies.

Practical Implications

International market researchers using samples including respondents from more than one cultural background have to be aware of the potential source of misinterpretation caused by cross-cultural differences in response patterns. A simple ANOVA-based procedure allows researchers to determine whether data can be used in its uncorrected form.

Originality of the paper

We investigate cross-cultural response styles for new groups of respondents (Australian versus Asian), we extend the study from the investigation of extreme values to full response patterns, and we give market researchers in the international marketing context an indication of how high the level of potential misinterpretation can be and present a simple means of checking how necessary it is to account for cross-cultural differences in response behaviour.

Keywords: Cross-cultural response styles, rating scales, ordinal scales, Likert scales

Classification: Research paper

INTRODUCTION

Marketing research results can only be as good as the data they are based on. Marketing research in the international context is burdened additionally by the fact that data sets collected to analyse market structure of any kind typically contain responses from people with different cultural backgrounds. This could distort results and cause misinterpretations of data due to cross-cultural response styles. A realistic practical example would be market research in the tourism industry. A country with rich cultural heritage which positions itself as a major destination for cultural tourism may conduct a survey among tourists or potential tourists from different countries of origin in order to identify the best target country for an advertising campaign. Respondents state how important cultural heritage, sightseeing and other related aspects of their trip are when they choose a vacation destination using a typical seven-point Likert scale as the answer format. The analysis of results indicates that the USA is most suitable because the average importance values given by American respondents are the highest. Consequently, the national tourism organisation launches a major advertising campaign promoting their country and its cultural attractions to Americans. Possibly the conclusions tourism managers drew were flawed and merely reflected a tendency of American respondents to use the endpoints of the answer scale more to express their interest than, for instance, Chinese respondents did. If this is the case it is possible that a sub-optimal target country was chosen for the advertising campaign.

Cross-cultural studies are in danger drawing wrong conclusions from empirical data if respondents from different cultural backgrounds are included who demonstrate systematically different response patterns which are not content-related. If differences in response patterns are not accounted for (in the way that the question is formulated to minimize the occurrence of content-unrelated response patterns or by correcting for known patterns of response among respondents of different cultural background) a systematic measurement error is introduced in

the data leading to biased results, particularly when direct comparisons between cultural groups are undertaken on the basis of such data.

Paulhus (1991, p.17) defines a *response bias* 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)". In addition he claims that "To the extent that an individual displays the bias consistently across time and situations, the bias is said to be a *response style*". The two most common forms of response styles are referred to as the Extreme Response Style (ERS) and the Acquiescence Response Style (ARS). ERS manifests itself by a tendency of respondents to use the endpoints of the answer scale in the questionnaire. The tourism example above illustrates an ERS. ARS describes the tendency of respondents to agree with items despite having a range of possibilities. In the above tourism example ARS could also be a reason for managerial misinterpretation of results if, for instance, the Japanese respondents would tend to strongly agree with the culture-related travel motivation questions. Again, Japan could be wrongly chosen as a suitable target market for advertising based on the failure to account for the fact that the respondents of Japanese tourists may be affected by ARS.

Technically, ERS increases reliability and decreases validity (Clarke III, 2001). It skews the frequency distribution to the ends, which increases standard deviation and decreases correlation, i.e. all correlation based methods, such as factor or regression analysis, are affected (Chun, Cambell & Yoo, 1974; Hui & Triandis, 1989; Heide & Gronhaug, 1992). ARS distorts regression analysis and might manifest itself in a spurious factor exclusively composed of negatively-keyed variables (Heide & Gronhaug, 1992). The presence of respondents tending to prefer the upper or lower end of the scale gives spuriously higher correlations and thus covariance-based analyses can be substantially influenced (Rossi, Gilula

& Allenby, 2001). Furthermore, segmentation results based on distances will be distorted, while the results will be the same if correlations are used (Greenleaf, 1992a).

Numerous previous studies have investigated the existence of differences in such cross cultural response styles, using questionnaires of different content and comparing respondents from different cultural groups. Most research into cross-cultural response styles, however, has used frequencies of ticked endpoints on the answer scale to evaluate response styles and compare them. Also, the extent of potential data distortion has, to the authors' knowledge, so far not been quantified. However, this is essential in order to assess the danger of potential misinterpretation if differences in cross-cultural response styles are not investigated before market structure analyses are conducted.

We extend past work in two directions:

(1) We investigate response patterns, thus investigating systematic differences in responses along the entire answer scale, not only the endpoints. The underlying assumption is that response styles do not only manifest themselves at endpoints. However, including the values along the entire answer scale should still permit accurate discrimination between respondents who prefer using endpoints and those who tend to tick middle categories. In particular, we investigate two hypotheses: (H1) The occurrence of extreme response styles is lower among Asian respondents, and (H2) Asian respondents demonstrate significantly different response patterns compared to Australian respondents.

(2) We quantify the extent of data distortion attributable to systematic differences in cross-cultural response patterns.

The article is structured as follows. We first review relevant past studies in the area of differences in cross-cultural response styles. We then describe the data sets used in the study. The results section is subdivided into three parts. The first part investigates whether the findings on differences in extreme response style reported in previous studies can be

replicated for the comparison of Australian and Asian respondents using the traditional method of comparing endpoint use. The second part investigates whether findings on differences in ERS reported in previous studies can be extended to response patterns including all answer options. The third part assesses the extent to which differences between cultures with respect to response patterns contaminates the data sets. Finally, we draw conclusions, and provide an overview of available correction techniques for data contaminated by response styles. We then discuss limitations and directions for future work.

PRIOR WORK

Extensive research work has been conducted in the past aiming to determine the influence of socio-demographic variables such as gender, age, or nationality on the individuals' response styles. While many have found that gender generally is not to be associated with response styles (Bachman & O'Malley, 1984; Watkins & Cheung, 1995; Clarke III, 2000; 2001), most studies - with the exception of Marshall and Lee (1998) - indicate that there are significant cultural differences. Such differences between countries within the European Union have been analysed by van Herk, Poortinga and Verhallen (2004) and Welkenhuysen-Gybels, Billiet and Cambre (2003), between Hispanic and Non-Hispanic respondents by Hui and Triandis (1989) and Marin, Gamba, and Marin (1992), and between Western and Asian countries by Zax and Takahashi (1967), Chun et al. (1974), Marshall and Lee (1998), and Roster, Rogers and Albaum (2003). With respect to the difference between Asian and Western countries, Chun et al. (1974) conclude that American respondents demonstrate higher ERS scores than Korean participants in surveys. Roster et al. (2003) show that the Chinese sample in their study demonstrated the lowest ERS scores. Contrarily, Zax and Takahashi (1967) find that Japanese female students respond in a more extreme manner than their American counterparts.

Explanations for cultural differences are given by Hui and Triandis (1989) and Stening and Everett (1984). Hui and Triandis argue that the matching of the underlying continuous construct to the categories offered varies between cultures. Another reason for cultural differences might be that cultural values, which influence the answer behaviour, are not the same. Stening and Everett argue that it is more important in Asian cultures to be modest and respond cautiously, while Hispanics associate sincerity with the use of the endpoints. Clarke III (2001) has provided a review of studies on difference in cross-cultural response styles. Such differences in response styles have typically been detected when ordinal answer formats such as Likert scales were used, which researchers have repeatedly found to be susceptible to response styles. This fact and the gravity of the problem have been known for half a century, leading to recommendations as drastic as the use of dichotomous scales in order to avoid data contamination caused by response styles (Cronbach, 1950). Recently, Baumgartner and Steenkamp (2001) published a systematics of response styles, distinguishing no less than seven: acquiescence response style (ARS), disacquiescence response style (DARS), net acquiescence response style (NARS), extreme response style (ERS), response range (RR), midpoint responding (MPR), and noncontingent responding (NCR). Despite this huge potential for data distortions due to systematic answer tendencies in respondents, most research studies on response styles conducted in the past focus on ERS (for instance, Chun, et al, 1974; Greenleaf, 1992b; Marshall & Lee, 1998; Clarke III, 2001). Only few studies were devoted to or included ARS and DARS (Greenleaf, 1992a; Marin et al., 1992; Cheung & Rensvold, 2000; Clarke III, 2001).

Generally, ERS is measured by counting the relative number of endpoints ticked or by the standard deviation of the answers, which according to Baumgartner and Steenkamp (2001) is in fact an appropriate measure for RR. Greenleaf (1992b) argues that the relative number of endpoints ticked is only a valid measure for ERS if (1) the inter-item correlation is

small and (2) the extreme response probability is similar for all items. To address measurement deficiencies caused by simply counting endpoints ticked, he proposes the use of a scale specifically developed to measure each respondent's ERS. Furthermore, he suggests that the correlations between the split-half ERS scores should be benchmarked against the expected value of a fitted beta-binomial model to verify the validity of the measure. He argues that the ERS scores theoretically follow a beta-binomial distribution, because in general the probability of ticking an endpoint is not constant over the population, but can be modelled by a beta distribution, thus accounting for heterogeneity in the population. While Greenleaf's method of measuring ERS is the most cited procedure, a number of other researchers have made alternative recommendations regarding how to identify and/or correct for response styles in ordinal data sets (Cunningham, Cunningham & Green, 1977; Heide & Gronhaug, 1992; Watson, 1992; Chang, 1994; Rossi et al., 2001; Van de Vijver & Poortinga, 2002; Welkenhuysen-Gybels et al., 2003).

The differences in extreme response styles have been studied using different multi-category scales ranging from 3-point to 10-point scales, where most studies used only one single answer format. Four-point scales were used by Marin et al. (1992) and Byrne and Campbell (1999), 5-point scales by Chun et al. (1974), Hui and Triandis (1989), Marin et al. (1992), Marshall and Lee (1998), and Baumgartner and Steenkamp (2001), 6-point scales by Greenleaf (1992a, 1992b) and 10-point scales by Hui and Triandis (1989) and Roster et al. (2003).

Despite Cronbach's early recommendation that researchers use fewer scale categories to avoid response style effects, few studies have examined the effect of the number of response categories offered on the differences in cross-cultural response style (Hui and Triandis, 1989; Clarke III, 2000; 2001). Hui and Triandis (1989) find that the number of extreme responses is comparable between Hispanic and Non-Hispanic for the 10-point scale,

while they are not for the 5-point scale. By contrast Clarke III (2000, 2001) finds that using a higher number of scale options is more susceptible to culturally determined response styles, thus presenting an interesting trade-off situation for the selection of response scales: is it better to accept higher levels of extreme answers and by doing so avoid response style differences between cultures or is it preferable to offer many scale points, leading to less extreme answer tendencies but higher levels of contamination of data by culturally determined response styles requiring corrective measures?

DATA

Two data sets were used for this analysis. The **soft drinks data set** resulted from a survey of the general population. Respondents were intercepted at the mall and asked if they would be prepared to participate in a short survey. The questionnaire included six brands and nine attributes. The categories were labelled “strongly agree”, “agree”, “slightly agree”, “slightly disagree”, “disagree” and “strongly disagree”. The respondents were asked to express the level of association between brands and attributes by ticking the corresponding box. Boxes were vertically aligned. 252 usable responses are available from this survey. The **fast food data set** was collected at the University of Wollongong. The survey was conducted among students of an undergraduate marketing research subject during their tutorials. Respondents were asked to state to what extent they agreed that certain attributes described fast food chain brands well or not using a six-point Likert scale. The data set includes six brands and 11 attributes. The categories were labelled “perfectly applies”, “applies well”, “applies a little bit”, “does not particularly apply”, “does not describe well”, “does absolutely not describe” and the respondents were asked to fill in the number, which indicated their agreement with the statements. 131 usable responses are available from this survey.

Both product categories emerged from pre-studies at the exploratory research phase during which the aim was to identify product categories meeting two criteria: (1) they should be product categories relevant to respondents to avoid sacrifice in data quality resulting from brands that respondents are not familiar with and (2) there should be no a priori reason to believe that either Asian and Australian respondents would have different evaluations of the brands within the chosen product categories. This was important in order to avoid distorted responses arising from the need for respondents to evaluate irrelevant product categories.

The respondents in both surveys were asked to indicate their nationality. All respondents who identified themselves as either Australian or from an Asian country were included for analysis. It should be noted at this point that the classification into "Australian" and "Asian" is a reasonable – not perfect - approach for the purpose of the present study. Although it is an established approach among researchers in the field of cross-cultural studies who investigate differences between Asian and other cultures with respect to psychographic variables, such as values, self-esteem, etc. (see, for example, Kim**,** Li and Ng, 2005; Kim and Omizo, 2005) to include respondents from various Asian countries and subsume them as "Asian", there may well be systematic differences in response styles between respondents from different Asian countries. The same may be true for "Australian" respondents, given that a large proportion of Australian residents have migrated and may well have been raised in a different cultural setting.

Table I gives the relative number of respondents from each culture separately for each data set. Furthermore, the respondent characteristics with respect to gender and age are reported. For the fast food data set specific, information on the Asian country is available: 18 respondents were Chinese, four Indonesian, three Indian, two Malaysian and Thai and one Lebanese, Singaporean and Sri Lankan. For the soft drink data set only the continent of origin of the respondent was recorded, but respondents were also asked to indicate their first

language. The information about the first language was used to conduct a validity check of the nationality information used for grouping. 186 Australians (96%) state that English is their first language, the remaining respondents state alternative first languages. Among the 36 Asian respondents 64% stated that Mandarin / Cantonese is their first language, ten ticked "Other", seven "English", two "Arabic" and one "Thai". These results suggest that the nationality information is valid and can be used for grouping the respondents into two sub-groups: Asian and Australian respondents, with the majority of the Asian respondents likely to be Chinese.

---------- Take in Table I ----------

Any result emerging from the investigation of answer format effects can be questioned by stating that the differences in answers were wrongly attributed to the response style and that in fact systematic variation in people's expressed opinions (content) can explain detected differences. For instance, German respondents are known to express satisfaction levels with their tourism experiences in Austria (if they are asked to evaluate their vacation experience in relation to their expectations; TourMIS, 2006) that are systematically lower than those of visitors from other countries of origin. This is not, however, due to difference in how respondents answer the questions. The reason for these differences lies in the Germans' higher levels of prior experience. German tourists know what to expect in Austria and thus express both lower levels of positive surprise and lower levels of disappointment. In this study we have dealt with this problem by selecting global brand names as stimuli rather than using objects for evaluation that Australian and Asian respondents would be expected to respond to differently in terms of previous levels of awareness or usage. To evaluate whether or not this assumption is reasonable we have tested for usage differences among Australian

and Asian consumers of soft drinks. Respondents were asked to state how many glasses of soft drink they consume on average per week. Results from the analysis of variance indicate that no significant differences can be detected with respect to usage ($F = 0.25$, p-value = 0.62), supporting the assumption that there is no reason to assume a systematic difference in answer content between Asian and Australian respondents. For the fast food study, this product category emerged as particularly suitable from the exploratory research phase because students from all nationalities felt confident in evaluating fast food brands and because the level of interest in this product category was high among students. This is important in order to avoid distortions of responses resulting from answering questions of little personal relevance. Furthermore, reports from ACNielsen on the usage of fast food indicates that a number of Asian countries (Hong Kong, Malaysia, Philippines, Singapore, Thailand and China) have higher consumption levels for fast food than Australia and that McDonalds was the most popular with "68% of people in the 28 major countries picking it as the first choice […] KFC ranks No. 2 at 32% and Pizza Hut ranks No. 3 at 29%." (ACNielsen, 2004). This indicates that fast food chains can indeed be viewed as global brands equally familiar to respondents from both Australia and Asian countries. Market research results with respect to soft drinks are similar. While the US is still the biggest market, it is expected that Asia and Australia will outpace the US in 2007, with an expected market share of 25% as opposed to 24% in the US. China is in fact already ranked among the top three markets for carbonates (Zenith, 2002). Given the test results for the soft drinks data, the selection mechanism for the fast food category and the market research figures, it is reasonable to assume that Asian and Australian respondents are equally familiar with fast food chains and soft drink brands. Furthermore, we used the same multi-category scale format (six-point) in both surveys to avoid any additional answer format effects arising from minor format variations.

The proportion of Asian respondents in the soft drink data set is higher than in the city population (12%), while the proportion of Asian students in the fast food data set is very well represented in the sample (27% of undergraduate students in the Commerce Faculty are from Asian countries). However, representativeness is in fact not needed to investigate the research questions under study. This study compares two subgroups of population and no conclusions are drawn with respect to the proportion of people in the respective populations. The sample sizes are small, but sufficient to conduct all the tests required for analysis and to indicate that there are significant differences between Asian and Australian respondents. Hence lack of power which is a known problem for small sample sizes is not of concern in this case.

The fact that the fast food data set uses a student convenience sample was not expected to impact negatively on the study results. Although students are clearly not representative of the general population in terms of socio-demographics or even the evaluation of fast food brands, they do represent a legitimate subset of the population with respect to the behaviour under study: the tendency to use answer formats in certain ways. There is no plausible hypothesis indicating why students should be more or less likely to be using different response styles than other groups.

Please note that the two samples are analysed independently. The reason for including two samples was to increase the external validity of conclusions derived from this study by investigating the research question in different settings and in different populations.

ANALYSIS AND RESULTS

All computations and graphics for the analysis of the empirical data have been developed using the R statistical software package (R Development Core Team, 2004).

ERS differs across respondents from different countries of origin

The cultural / national differences with respect to the use of extreme response categories were investigated as suggested by Roster et al. (2003). The internal reliability of ERS scores was checked using split-half methods (Chun et al., 1974; Greenleaf, 1992b).

The first analysis of the difference with respect to ERS across the two cultures is conducted using the endpoint-approach discussed above. The relative number of extreme endpoints ticked is counted. In order to validate the resulting ERS scores, their internal reliability is investigated: first, answers of each respondent are randomly split into two halves. Then, the two separate ERS scores are correlated. This correlation is benchmarked against the reliability derived, assuming that the scores are generated from an underlying beta-binomial model. In order to get a more reliable estimate of the correlation between the split-half ERS scores, we report the mean of 10 random splittings. The parameters of the beta-binomial distribution are determined separately for each culture using the moment estimation procedure proposed in Greenleaf (1992b). The results are given in Table II. The predicted and observed internal reliability coefficients are higher if the number of endpoints ticked is higher. They are therefore lowest for the negative endpoints for the Asian respondents for the soft drink data set. These internal reliability coefficients are about 50, whereas the remaining coefficients are between 70 and 85. The internal reliability coefficients indicate that the ERS scores represent a valid measure of that response style because the estimated reliability scores correspond to those predicted by the beta-binomial model. Furthermore, there is no evidence that the ERS scores are more reliable for one of the two cultures.

---------- Take in Table II ----------

We investigated the difference in ERS between cultures by comparing the relative number of extreme negative and/or positive responses averaged for each of the two cultures. The results for the two data sets are presented in Table III. The use of endpoints is rather similar for the two data sets, even though they are slightly more frequently ticked for the fast food data set. However, the separate analysis of the endpoints reveals that there are strong differences as the negative endpoint is only very infrequently ticked for the soft drink data set. Comparison between the cultures indicates that the endpoints are less frequently ticked by the Asian than by the Australian respondents. We tested the significance of the differences using a two-sample test for equality of proportions with continuity correction. For each comparison a significant difference is indicated between the Asian and Australian respondents as the p-values are all smaller than 0.04.

----------Take in Table III ----------

In both data sets we found significant differences between Australian and Asian respondents. In addition, the table shows that the latter use endpoints less frequently. This is true whether analysing the positive and negative endpoints alone or if the endpoints are studied in an aggregated matter. These findings support our hypothesis (H1) that the occurrence of extreme response styles is lower among Asian respondents.

Response patterns differ across respondents from different countries of origin

Different response styles lead to different probabilities for the categories to be ticked by a respondent regardless of the content. Respondents with an extreme response style are more likely to tick endpoints, while respondents with a mild response style have a lower probability of doing so. Answer patterns can be determined for each respondent by using the relative frequency of each category given the respondent's answers for the complete

questionnaire. In general, the estimated answer patterns reflect not only the probability of ticking a category but also the attitude of the respondents towards the complete questionnaire (e.g. the overall interest and involvement with fast food chains) and the characteristics of the questionnaire (e.g. if a balanced wording scheme is used). Answer patterns therefore not only capture potential response bias introduced to a cross-cultural comparison by response styles, but also reflect the overall differences in attitude and the characteristics of the questionnaires. Detected differences in response patterns can therefore not be unambiguously attributed to response bias introduced by response styles. This makes a correction difficult, as this requires the assumption that differences are solely attributable to response styles and not to overall attitudes.

Because of this influence of the characteristics of the questionnaire, we performed a separate analysis of the answer patterns for each data set. The respondents did not always answer all questions and the answer patterns are based for each respondent on average on 65.7 answers out of 66 for the fast food data set, and on 53.3 answers out of 54 for the soft drink data set.

The following analysis is based on the assumption that respondents with extreme response styles will use endpoints more frequently, respondents with mild response styles will make more use of the middle range of the scale and other respondents will evenly distribute their answers across all answer scale points. In the latter cases, their response styles will be referred to as “even”. In general the response styles of individuals from a population of respondents will vary continually from extreme to mild. A partition of the respondents into three segments – each representing one response style – might be useful for determining the strength of these differences. We achieved this by analysing the centroids resulting from a grouping procedure. In addition, the size of the resulting response style segments provides an

indication of how strongly each response style is represented among respondents of each of the data sets.

Respondents' answer patterns were partitioned using the *k*-means algorithm (Hartigan and Wong, 1979). This is an iterative grouping procedure that aims at minimising the difference between respondents assigned to the same group while maximising the difference between groups. In order to avoid a suboptimal grouping due to local optima, we used 100 different random initializations. The optimal solution with respect to within-cluster sum of squares is reported. The centroids for both data sets are given in Figure 1, labelled by the response style they represent. These centroids reflect the answer patterns that would be theoretically expected to emerge very well. Figure 1 also indicates that the results are very similar (both indicating the existence of a mild, extreme and even answer pattern) for the data sets collected on campus with student respondents and at shopping malls using the general population. These results empirically support the above argument that the same response styles exist in subgroups of the total population. The difference in the distribution across answer options, evident from Figure 1, is due to the fact that the fast food data included a higher number of negative attributes (six positive attributes, such as yummy and cheap, four negative attributes, such as greasy and fattening and one attribute – spicy - which cannot generally be classified as either positive or negative) than the soft drink questionnaire contained (two positive attributes, such as refreshing, one negative attribute – unhealthy – and five attributes which can not be generally classified as either positive or negative, such as sweet). This difference in the nature of items leads to more symmetrical answer patterns in the fast food data as opposed to the soft drink data.

---------- Figure 1 ----------

While the answer patterns of respondents (centroids of the segments) are very similar to each other, the overall segment sizes differ. The mild segment for the fast food is very distinct, with virtually no extreme negative statements. It is consequently not surprising that the size of this group of respondents is fairly small, including only 5 percent of the students. Contrarily, the extreme segment contains more than twice as many members when fast food chain brands were evaluated than is the case when soft drink brand images are under study. The relative segment sizes for each of the data sets are given in Table IV.

---------- Take in Table IV ----------

Table 4 also includes the relative segment sizes for each of the cultures. As can be seen, Asian respondents are more likely to have a mild response while Australians are more likely to use extreme values. This difference is significant for both data sets as indicated by Fisher's exact test for count data (Fast food: p-value = 0.007; Soft drink: p-value = 0.003). The association between the segmentation and culture is further illustrated in Figure 2 using mosaic plots where the vertical axis represents the three response styles and the horizontal axis shows the culture of the respondents for each of the data sets. The area of each rectangle represents the relative size of each response style segment-culture combination separately for each data set. The shading of the rectangles indicates the significance of the residuals of a log-linear model, assuming independence in each cell of the cross-tabulation for a given data set. The significance of the shading codes is indicated on the right of the figure.

---------- Figure 2 ----------

The figure indicates that the significant association between culture and segment membership is due to the over-representation of Asian respondents in the mild segment. These findings support our hypothesis (H2) asserting that Asian respondents demonstrate significantly different response patterns to Australian respondents. More specifically – and in line with the findings related to hypothesis (H1) – Asian respondents are more likely than Australian respondents to be members of the mild segment.

Extent of data contamination through differences in cross-cultural response patterns

In the following we analyse the amount of variability in the response patterns explained by culture and the response style segments. The inter- and intra-cultural variability is determined by an analysis of variance (ANOVA) using the sum of squares (SS) of the answer patterns aggregated over all answer categories. The overall SS are split into the SS between cultures, within cultures and between segments of the *k*-means partitions and within cultures and segments. The SS between cultures indicates how much variability in the response patterns can be attributed to culture. The SS within cultures and between segments shows how much variability is explained by a suitable categorization of the respondents with respect to their response styles. In Table V we present the mean SS, together with the relative amount of variability explained by each SS.

---------- Take in Table V ----------

The mean squares are higher for values between segments and within cultures than between cultures. They are lowest for those within segments and cultures, representing the residual mean squares after accounting for culture and the response style segments suggested by *k*-means. This variability can be seen to include the individual heterogeneity in the response patterns after eliminating noise attributable to culture or response style segment.

The variability explainable by culture is rather small, with 6% for the fast food data set and 2% for the soft drink data set. Nevertheless, the mean SS between cultures are highly significant if compared to the mean SS within cultures (Fast food: F-value=8.1, Soft drink: F-value =5.3– all p-values < 0.001).

The SS between segments and within cultures is determined by summing up – over all respondents - the squared differences between the mean values of the respondents with the same segment membership and culture and the mean values of the respondents with the same culture. This SS explains about 40% of the variability. The SS between segments without using any information on culture was maximized by *k*-means.

The analysis of the SS indicates that data aggregated by culture has to be corrected for the different use of the ordinal scales. However, if the data is analysed with respect to the answers of the individuals, it is even more important to account for intra-cultural (i.e. individual) differences as only a small amount of the variability in the answer pattern is found to be explained by culture.

The analysis of the SS for the response patterns indicates that there are small but significant differences between cultures. As it is assumed that the patterns are primarily determined by the response styles of the respondents, it can be concluded that it is necessary to account for differences in culture in order to be able to make pancultural analyses.

CONCLUSIONS AND FUTURE RESEARCH

The results regarding cross-cultural response styles based on the analysis of ERS replicate prior findings that cultural differences in response styles exist. The comparison of Australian and Asian respondents indicates that the latter are more likely to exhibit a mild response style leading to less extreme points ticked and lower ERS scores. The reliability of

the ERS scores was comparable for both cultures and corresponded to the hypothesized values of a beta-binomial model as suggested by Greenleaf (1992b).

In addition to an analysis using only the endpoints, the answer patterns that reflect the use of the categories of the complete scale were used to determine segments of respondents with similar response styles. This method effectively separated the respondents into segments corresponding to a mild, even or extreme response style.

The analysis of the answer patterns using SS indicates that there are significant differences between cultures, even though the amount of heterogeneity within cultures is also substantial. Culture and the partition into three response style segments accounted for about half of the variability in the answer patterns. The inter-cultural variability accounted for between 6 to 2 percent of variability, whereas the segmentation covered about 40 percent.

The significant differences between cultures suggest that aggregated data cannot be compared (for instance, in a statement such as “More Australian respondents think that bright colours are important when buying a sports shoe than Asian respondents”, leading to potential managerial decisions to modify the colour codes of sports shoes for the Australian market) without correcting for differences in response styles to assure that comparisons are based on the content component of the response only.

Difference between cultures can be easily checked by performing an ANOVA on the answer patterns given the different cultures. If the difference is not significant the data can be safely combined. If however, this difference is significant, the researcher has to decide whether the differences encountered are meaningful findings in terms of cross-cultural differences or whether they are likely to represent differences in response behaviour, in which case the original data set should be corrected before analysis. An overview on possible correction techniques is provided below. In order to minimize the difference in answer patterns the use of scales less susceptible to scale usage heterogeneity (for instance, binary or

3-point scales) could be considered in the development of survey tools for cross-cultural studies. Baumgartner and Steenkamp (2001) recommend the use of balanced scales as this attenuates the influence of ARS.

If it is assumed that findings might be distorted due to systematic differences in response styles, it is necessary to account for these differences before analysing the data. Standardization methods are the most popular in this context, because no specific scales or predictive variables are needed and they are easy to apply. In addition, model-based approaches to account for scale usage heterogeneity are considered in Rossi et al. (2001), Wolfe and Firth (2002) and Johnson (2003).

Fischer (2004) gives an overview on the different standardization methods proposed with respect to different units and forms of adjustments. Units of adjustments might be subjects, groups (i.e. variables or items), cultures or a combination of them depending on which level the difference in response styles is assumed to reflect and over which level comparability is desired. The different forms of adjustments use either means, dispersion indices, both or covariates depending on which kind of response style shall be removed. We propose using means, in general, for accounting for ARS, while we are dividing through dispersion indices such as standard deviations to remove ERS. We base standardization on the assumption that response styles are constant within the unit and that it is possible to capture them using the observations within each unit. However, through standardization content-related differences might also be lost. A further disadvantage is that subtracting the mean leads to ipsatized, scales which are known to reflect only intra-unit (relative) differences (Chan 2003) and to distort correlation estimates if inter-item correlations are high (Bartram 1996). These assumptions and disadvantages have to be regarded if a specific standardization method is selected by choosing the suitable unit and form of adjustment.

Model-based approaches assume that a latent, continuous variable is measured on an ordinal scale and the mapping between the latent variable and the answer point ticked is influenced by the response style of the respondent. Rossi et al. (2001) show that their approach outperforms standardization methods as described above. However, their correction technique is based on the assumption that respondents are homogeneous in their responses with respect to the latent variable and that differences are solely contributable to response styles. This assumption might not be suitable in certain applications, e.g. if the aim of the analysis is to perform a market segmentation.

A limitation of this study is that respondents from different Asian countries and all respondents classifying themselves as Australian were combined under the assumption that they have a common cultural background. In the future it would be interesting to further investigate the homogeneity of respondents from different Asian countries with respect to response styles and account for the cultural group Australia respondents identify with most. It is expected that such an approach would further increase the amount of variability in the response patterns explained by culture. In addition it would be interesting to include a comparison to another cultural group, for example Hispanic respondents, who have been shown to tend to a more extreme response style. Furthermore, only one answer format was analysed. The methods proposed should be applied to surveys with other answer formats in order to investigate their suitability for other scales. Another direction of future research would be to include acculturation into the investigation and assess whether there is an acculturation effect in how respondents react to answer formats. This analysis was not possible with the present data sets because no information on the level of acculturation was included in the survey and because the samples size would have been too small to further subdivide them into acculturation level subgroups within cultures.

ACKNOWLEDGEMENTS

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

Table I

Demographic characteristics of respondents

	Percentage distribution			
	Fast food		**Soft drink**	
Characteristic	Asian	Australian	Asian	Australian
Size	(N = 131)		(N = 252)	
	24.4	75.6	23.4	76.6
Gender				
Female	34.4	36.1	59.3	47.2
Male	65.6	63.9	40.7	52.8
Age				
Less than 26			83.0	61.6
26-35			9.4	21.5
36-50			7.6	13.0
51 and above			0.0	4.0

Table II

Internal reliability of ERS scores

	Internal reliability coefficient			
	Asian	**Model Asian**	**Australian**	**Model Australian**
Fast food				
Both	87.8	86.0	81.2	78.3
Positive	83.4	77.4	76.6	68.6
Negative	81.0	75.9	75.3	68.0
Soft drink				
Both	81.4	81.5	86.3	85.8
Positive	80.0	80.4	83.0	81.6
Negative	54.7	46.9	82.8	84.0

Table III

Between-culture use of extreme scale positions

	Average use of extreme scale positions for		
	Both	**Positive**	**Negative**
Fast food			
Asian	22.2	13.2	9.0
Australian	36.0	21.2	14.8
X^2	138.0	65.1	45.9
Significance of Difference	< 0.001	< 0.001	< 0.001
Soft drink			
Asian	20.9	18.8	2.1
Australian	23.8	20.5	3.2
X^2	10.9	4.4	10.4
Significance of Difference	< 0.001	0.037	0.001

Table IV

Segment sizes

	Percentage of respondents in each of the segments		
	mild	**even**	**extreme**
Fast food			
Overall	5.3	48.1	46.6
Asian	15.6	53.1	31.2
Australian	2.0	46.5	51.5
Soft drink			
Overall	40.1	38.9	21.0
Asian	59.3	28.8	11.9
Australian	34.2	42.0	23.8

Table V

Inter- and intra-cultural differences for answer patterns

	Fast food		Soft drink	
	Mean Sq.	**Rel. SS**	**Mean Sq.**	**Rel. SS**
Between cultures	0.07	0.06	0.08	0.02
Between segments and within cultures	0.12	0.40	0.44	0.44
Within segments and culture	0.01	0.55	0.01	0.54

FIGURES

Figure 1

Segmentation of answer patterns on a 6-point scale

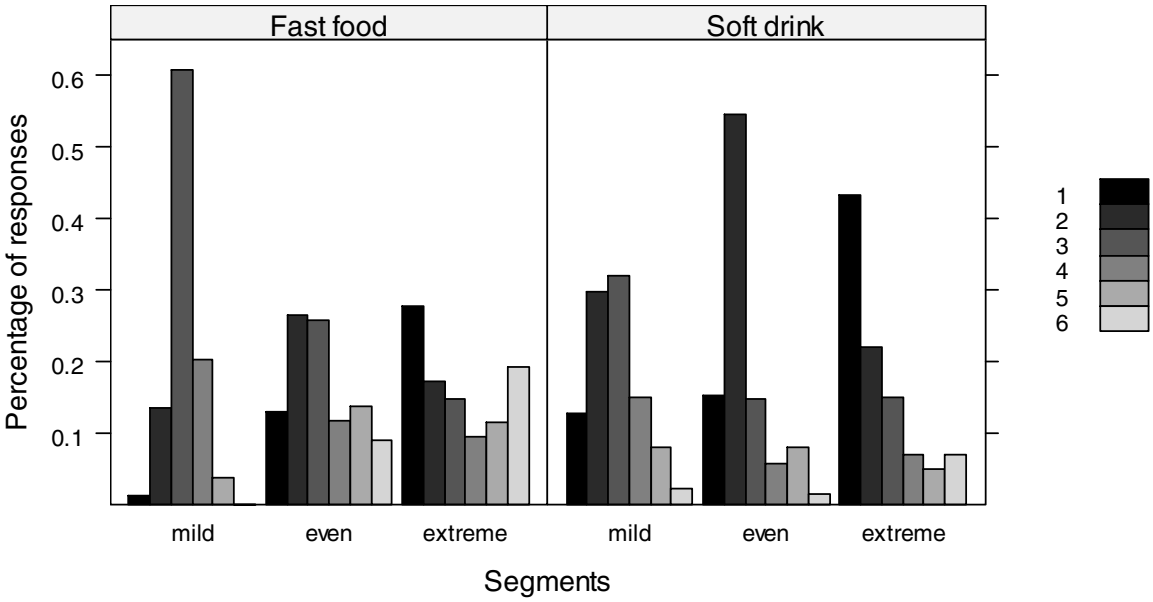
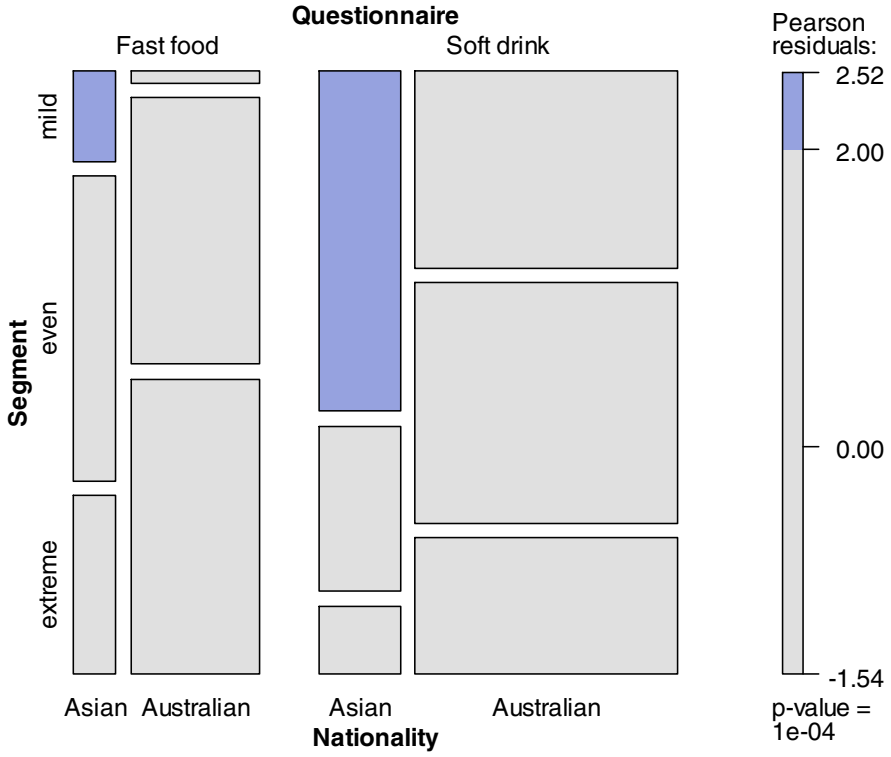


Figure 2

Association between response style segments and culture on a 6-point scale



REFERENCES

ACNielsen (2004), "Consumers in Asia Pacific – Our Fast Food / Take Away Consumption Habits", available at http://www2.acnielsen.com/reports/, last accessed 15.8.2006.

Bachman, J. G., & O'Malley, P. M. (1984), "Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles", *Public Opinion Quarterly*, vol.48 no. 2, pp.491-509.

Baumgartner, H., & Steenkamp, J. (2001), "Response styles in marketing research: A cross-national investigation", *Journal of Marketing Research*, vol. 38 no. 2, pp.143-156.

Bartram, D. (1996), "The relationship between ipsatized and normative measures of personality", *Journal of Occupational and Organizational Psychology* vol. 69, pp.25-39.

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*, vol. 30 no. 5, pp.555-574.

Chan, W. (2003), "Analyzing ipsative data in psychological research", *Behaviormetrika*, vol. 30 no. 1, pp.99-121.

Chang, L. (1994), "A psychometric evaluation of 4-Point and 6-Point Likert-type scales in relation to reliability and validity", *Applied Psychological Measurement*, vol. 18 no.3, pp.205-215.

Cheung, G. W., & Rensvold, R. B. (2000), "Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling", *Journal of Cross-Cultural Psychology*, vol. 31 no. 2, pp.187-212.

Chun, K. T., Campbell, J. B., & Yoo, J. H. (1974), "Extreme response style in cross-cultural research – Reminder", *Journal of Cross-Cultural Psychology*, vol. 5 no. 4, pp.465-480.

Clarke III, I. (2000), "Extreme response style in cross cultural research: An empirical investigation", *Journal of Social Behaviour and Personality*, vol. 15 no.1, pp.137-152.

Clarke III, I. (2001), "Extreme response style in cross-cultural research", *International Marketing Review*, vol. 18 no. 3, pp.301-324.

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

Cunningham, W. H., Cunningham, I. C., & Green, R. T. (1977), "The ipsative process to reduce response set bias", *The Public Opinion Quarterly*, vol. 41 no.3, pp.379-384.

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*, vol. 35 no.3, pp.263-282.

Greenleaf, E. A. (1992a), "Improving rating-scale measures by detecting and correcting bias components in some response styles", *Journal of Marketing Research*, vol. 29 no.2, pp.176-188.

Greenleaf, E. A. (1992b), "Measuring extreme response style", *Public Opinion Quarterly*, vol. 56 no.3, pp.328-351.

Hartigan, J.A. & Wong, M.A. (1979), "Algorithm AS 136: A K-means clustering algorithm", *Applied Statistics*, vol. 28 no.1, pp.100-108.

Heide, M., & Gronhaug, K. (1992), "The impact of response styles in surveys: A simulation study", *Journal of the Market Research Society*, vol. 34 no.3, pp.215-223.

Hui, C. H., & Triandis, H. C. (1989), "Effects of culture and response format on extreme response style", *Journal of Cross-Cultural Psychology*, vol. 20 no.3, pp.296-309.

Johnson, T.R. (2003), "On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style", *Psychometrika*, vol. 68 no.4, pp.563-583.

Kim, B. S. K., Li, L. C., & Ng, G. F. (2005), "The Asian American values scale - Multidimensional: Development, reliability, and validity". *Cultural Diversity and Ethnic Minority Psychology,* vol. 11, pp.187-201.

Kim, B. S. K., & Omizo, M. M. (2005), "Asian and European American cultural values, collective self-esteem, acculturative stress, cognitive flexibility, and general self-efficacy among Asian American college students". *Journal of Counseling Psychology*, vol.52, pp.412-419.

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,* vol. 23 no.4, pp.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, The Association for Consumer Research, pp. 90-95.

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, pp.17-59.

R Development Core Team (2004), R: A language and environment for statistical computing, R Foundation for Statistical Computing: Vienna.

Rossi, P.E., Gilula, Z. & Allenby, G.M. (2001), "Overcoming scale usage heterogeneity: A Bayesian hierarchical approach", *Journal of the American Statistical Association*, vol. 96 no. 453, pp.20-31.

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.

Stening, B. W., & Everett, J. E. (1984), "Response styles in a cross-cultural managerial study", *Journal of Social Psychology*, vol. 122 no.2, pp.151-156.

TourMIS (2006). Gaestebefragung Oesterreich [online]. Available: http://tourmis.wu-wien.ac.at/ [Accessed 2006, Feb. 17].

van de Vijver, F. J. R., & Poortinga, Y. H. (2002), "Structural equivalence in multilevel research", *Journal of Cross-Cultural Psychology*, vol. 33 no.2, pp.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*, vol. 35 no.3, pp.346-360.

Watkins, D., & Cheung, S. (1995), "Culture, gender and response bias", *Journal of Cross-Cultural Psychology*, vol. 26 no.5, pp.490-504.

Watson, D. (1992), "Correcting for acquiescent response bias in the absence of a balanced scale: An application to class consciousness", *Sociological Methods and Research*, vol. 21 no.1, pp.52-88.

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*, vol. 34 no.6, pp.702-722.

Wolfe, R., & Firth, D. (2002), "Modelling subjective use of an ordinal response scale in a many period crossover experiment", *Applied Statistics*, vol. 51 no.2, pp.245-255.

Zax, M., & Takahashi, S. (1967), “Cultural influences on response style: Comparison of Japanese and American college students”, *Journal of Social Psychology*, vol. 71, pp.2-10.

Zenith (2002) “Soft Drinks”, available at http://www.zenithinternational.com, last accessed 15.8.2006.