Response style differences in cross-national research: dispositional and situational determinants

Anne-Wil Harzing
Michelle Brown
Kathrin Köster
Shuming Zhao

Version January 2011

Prof. Anne-Wil Harzing   Email: anne-wil@harzing.com
University of Melbourne   Web: www.harzing.com
Department of Management & Marketing
Faculty of Economics & Commerce
Parkville Campus
Melbourne, VIC 3010
Australia

Response style differences in cross-national research:

dispositional and situational determinants*

Prof. Anne-Wil Harzing
Department of Management & Marketing, University of Melbourne, Parkville, Australia, Email:
harzing@unimelb.edu.au

Assoc. Prof. Michelle Brown, Department of Management & Marketing, University of Melbourne, Parkville, Australia

Prof. Kathrin Köster, Department of Business Administration II, Heilbronn University, Heilbronn, Germany

Prof. Shuming Zhao, School of Business, Nanjing University, Nanjing, China

Abstract

- International Management researchers often rely on surveys to collect their data. However, responses to survey questions can be biased by response styles, a respondent's tendency to provide a systematic response to questions regardless of their content. Response styles vary across countries and individuals, but there is limited systematic research that investigates *why* they vary.

- Our study investigates middle (MRS) versus extreme response styles (ERS), the tendency to use the middle or extreme categories on rating scales. We examine the impact of culture, different types of scale anchors and the level of knowledge of the topic in question on MRS and ERS.

- We asked five groups of respondents (Chinese in China, Chinese in Australia, Anglo-Australians in Australia, and two groups of German students in Germany) to indicate on

a 10-point scale whether certain employee attitudes or behaviour were more typically Australian (left-hand of the scale) or Chinese (right-hand of the scale). We then asked them how they would rate the performance (low to high on a 10-point scale) of an employee who displayed this attitude or behaviour.

- Asian respondents showed higher MRS than Western respondents. When scale anchors referred to naturally opposing and mutually exclusive constructs (Australian versus Chinese) respondents showed more ERS than when they referred to level or degree of a construct (low-high performance). Knowledge of cross national differences resulted in higher ERS on behavioural questions but not on performance questions.

Key words: response styles, survey research, culture, China, Australia, Germany

Abbreviated title: Response style differences in cross-national research

Introduction

Researchers in the field of International Management use a wide variety of research methods. Two of the most frequently used methods and data sources are survey research or the use of secondary data, in itself often based on surveys conducted in the past (see e.g. Buckley, Devinney & Louviere's (2007) review of studies on FDI location choice). Researchers would normally assume that responses to questionnaire surveys are only based on the substantive meaning of the questions involved. This might be true for surveys that deal with issues that can be measured objectively, such as for instance the classification of a subsidiary as greenfield establishment or acquisition. However, there are many types of questions where respondents may display a bias in their responses.

These biases could include a consistency bias, i.e. the respondents' desire to be seen as consistent by the researcher, which may lead them to respond in accordance with a presumed underlying relationship, for instance the assumption that certain practices should lead to high performance. It could also refer to social desirability bias, the tendency to respond in a way that is seen as socially desirable. This can be expected to be especially relevant in questions that are for instance dealing with ethics or corporate social responsibility. Researchers will also find that questions relating to firm performance, when measured subjectively with Likert scales using low/high or below/above average anchors, often lead most respondents to score their performance above average.

However, there is another type of bias – response style bias – that can be invoked with *any* types of surveys that contain Likert-scale questions, which ask respondents to rate their opinions and attitudes. These questions could relate to individual cultural values or norms, but also to the respondent's assessment of the company's international strategy or the subsidiary's role. Response style bias is rather unique in that it is not dependent on the content of the question. The term 'response style' refers to a respondent's tendency to provide a systematic response to questionnaire items regardless of their content (Baumgartner & Steenkamp 2001). The

most common response styles are acquiescence (ARS) or disacquiescence (DRS); that is, the tendency to agree or disagree with an item regardless of the content, and extreme response styles (ERS) versus middle response styles (MRS); that is, the tendency to use the extreme or middle response categories on ratings scales (Harzing 2006).

Response styles can create a range of problems for researchers. First, they contaminate observed responses, because they either inflate (ARS) or deflate (DRS) respondents' scores on the measurement instrument in question (Baumgartner & Steenkamp 2001). They can also increase (ERS) or constrict (MRS) the response range. These effects are particularly problematic in comparing groups of respondents from different countries, as these have been shown to differ substantially in typical response styles (see e.g. Harzing 2006 for a recent summary). An international management researcher might therefore draw the erroneous conclusion that particular groups of respondents are different on the phenomenon under investigation, while in reality the groups only differ in terms of their response styles. Second, response styles can also influence the conclusions drawn about the *relationship* between variables as they can inflate or deflate the correlation between respondents' scores on the various scales (Baumgartner & Steenkamp 2001).

Acquiescence and extreme responses styles are the most commonly studied type of response style and have been recognized for half a century (see e.g. Couch & Keniston 1960; Cronbach 1946). However, these studies have all focused on Likert-scale questions with response anchors reflecting an *intensity dimension*, such as levels of agreement or levels of importance. The first part of this study follows a different approach by using scale anchors that *prevent* an acquiescence bias by offering two opposing and mutually exclusive options, in our case Australian versus Chinese behaviour. We therefore do not consider acquiescence bias in our study. Instead, we study extreme response styles (ERS) versus middle response styles (MRS); that is, the tendency to use the extreme or middle response categories on ratings scales. Although previous studies have looked at ERS versus MRS (see below for a summary), they have

all focused on Likert-scale questions with disagree/agree scales. As detailed below, in our study we contrast two different types of scale anchors. This allows us to significantly expand our knowledge of response style differences and hence our ability to get more meaningful answers in cross-national research.

Contributions of our study

In our study we asked five groups of respondents (Chinese in China, Chinese in Australia, Anglo-Australians in Australia, and two groups of German students in Germany) to read 90 statements about employee attitudes or behaviour and indicate whether they thought these attitudes or behaviours were more typically Australian (left hand side of the 10-point scale) or Chinese (right-hand side of the 10-point scale). We then presented them with the same 90 statements and asked them how they would rate the performance of an employee who displayed this type of behaviour or attitude on a 10-point scale from low performance (=1) to high performance (=10).

This approach has three advantages over previous studies. First, the use of countries as scale anchors in the first part of the study and performance implications of typical behaviour in these countries in the second part of the study, provides a more focused assessment of ERS versus MRS. In contrast to the use of disagree/agree Likert scales, these scale anchors do not conflate these response styles with ARS/DRS. When using Likert scales with disagree/agree formats it is difficult to distinguish positive ERS (the tendency to use the positive extremes of the scale) from ARS (the tendency to agree, i.e. use the positive extremes of the scale). The same is true for negative ERS and DRS.

Second, our scale anchors are less likely to be vulnerable to translation problems. As Harzing (2006) notes, differential interpretation in different languages of equivalent scale anchors might be an important explanation for response styles. Although scale anchors might translate into appropriate local equivalents, the intensity associated with these equivalents might

be different from the original language. Since our scale anchors in the first part of the study refer to specific country names rather than abstract constructs such as trust or motivation, it is unlikely that our translations are misinterpreted. Although the scale anchors in the second part of the study referred to low/high performance, this was framed in a country context, with alternating statements of Chinese and Australian behaviour. Hence, any differences in intensity between languages would be cancelled out as they would apply to both "Australian" and "Chinese" statements.

Third, the use of a 10-point scale is relatively uncommon in studies dealing with response styles. Hui & Triandis (1989) recommend increasing the number of response categories as a technique to reduce ERS. The rationale behind this is that scales with finer gradations provide more opportunity for the respondent to reflect their true opinion, without having to rely on extreme answers. Hui & Triandis (1989) recommend scales with more categories as particularly appropriate for cross-cultural research.

Our study builds on earlier studies by providing a coherent rationale for differences in response styles. Most of the earlier studies on response styles provided either descriptive results or limited post-hoc rationales for differences. More important for international management researchers is *why* response styles differ between countries and groups. Reasons for differences in response style can be dispositional – that is, related to individual characteristics such as age, gender or personality – or situational – that is, related to situational characteristics such as the format of the response scale, the ambiguity of questions, or time pressure (Baumgartner & Steenkamp 2001).

In the context of cross-country differences in response styles, cultural differences would be a likely dispositional explanation. In this article, we mainly focus on collectivism versus individualism and argue that MRS is likely to be more common in collectivistic countries. We also include a situational variable that has not been studied before: the type of scale anchors used. We argue that if the scale anchors refer to *naturally opposing and mutually exclusive con-*

structs, respondents are more likely to provide extreme responses than when scale anchors refer to a *level or degree of a construct*, e.g. agreement, performance or importance. To the best of our knowledge there are no studies that have systematically varied the *type* of scale anchors with the same question set. Finally, we include a dispositional variable that has not been studied before: the level of knowledge a respondent has about the topic in question. We argue that a higher level of knowledge is likely to positively influence extreme response styles. We find confirmation for all our hypotheses indicating that both dispositional and situational determinants have a strong influence on response styles. Our final section draws conclusions for cross cultural research and suggest additional avenues of study in this field.

Conceptual framework and hypotheses

National/cultural background

A range of previous studies has found strong differences in response styles between countries (see e.g. Harzing 2006 for a recent overview). In particular, East Asian respondents have been shown to display a higher proportion of middle response in comparison to US and Canadian respondents who displayed more extreme response styles (Chen, Lee & Stevenson 1995; Chun, Campbell & Yoo 1974; Lee & Green 1991; Shiomi & Loo 1999; Takahashi, Ohara, Antonucci & Akiyama. 2002; Zax & Takahashi 1967). Watkins & Cheung (1995) and Dolnicar & Grün (2007) confirmed that this difference was also apparent between Australian and Asian (the majority of which were Chinese) respondents, whilst Si & Cullen (1998) found similar differences between East Asian and Western managers from the US, Germany and the UK when scales with explicit mid-points are used.

These earlier studies have typically tended to offer post hoc and limited rationales for differences in response styles. An exception is Chen et al. (1995) who showed a relationship between self-reported individualism and extreme response styles. Similarly, Johnson, Kulesa, Cho

& Shavitt (2005) and Harzing (2006) suggested that national-level collectivism might be related to middle (and acquiescent) responses styles, a relationship also found by Smith (2004) for acquiescent responses. Collectivist countries tend to be characterized by a preference for harmony and an avoidance of confrontations, leading to more conformity in behaviour, where opinions are predetermined by the in-group (Hofstede 2001). This tendency is reinforced by the fact that most collectivist countries are characterized by an indirect communication style, where the expression of strong opinions is avoided (Hall 1976). Confucian teachings (predominant in East Asian countries) reinforce this communication style by advising followers to keep themselves from extremes (Si & Cullen 1998).

In collectivist countries there will be a preference for middle or slightly positive responses as these are most likely to avoid confrontation and preserve harmony. This argument is consistent with earlier studies that suggested evasiveness (the desire not to reveal one's true opinion) as one reason for middle response styles (see Baumgarten & Steenkamp 2001). In individualist countries, characterized by a lower emphasis on conformity and harmony and an acceptance of confrontations, individual initiative is expected and speaking one's mind is appreciated (see e.g. Gao, Kao & Ting-Toomey 1998). This is reinforced by the more direct communication style found in most individualist countries. We would therefore expect a higher willingness to disagree or offer extreme opinions in individualist countries.

Another line of reasoning is put forward by Hamamura, Heine & Paulhus (2008). They suggest that a higher endorsement of dialectical thinking might be the underlying reason for East Asian respondents to choose middle responses. Peng & Nisbett (1991) define dialectical thinking as the tolerance for holding apparently contradictory beliefs. They found, for instance, that two arguments that Americans viewed as incompatible were often both endorsed by Chinese respondents, resulting in a lower proportion of extreme responses.

The empirical research therefore suggests that respondents from collectivist countries typically display more compressed response styles than respondents from individualist coun-

tries, that is, display a preference for middle responses over extreme responses. In the context of our study's operationalisation this leads to the following hypotheses:

Hypothesis 1a: East Asian (Chinese) respondents will display a significantly lower standard deviation when classifying behaviour as typically Chinese or Australian than Western (Anglo-Australian/German) respondents.

Hypothesis 1b: East Asian (Chinese) respondents will display a significantly smaller differentiation between typical Chinese and Australian behaviour than Western (Anglo-Australian/German) respondents.

Hypothesis 1c: East Asian (Chinese) respondents will display a significantly lower standard deviation in assessing performance of typical Chinese/Australian behaviours than Western (Anglo-Australian/German) respondents.

Hypothesis 1d: East Asian (Chinese) respondents will display a significantly smaller differentiation in assessing performance of typical Chinese and Australian behaviour than Western (Anglo-Australian/German) respondents.

Scale anchors: classification versus judgment

We investigate a new situational variable: the type of scale anchor used. When scale anchors refer to *naturally opposing and mutually exclusive constructs*, respondents are more likely to provide extreme responses than when scale anchors refer to a *level or degree of a construct*, e.g. agreement, performance or importance. Naturally opposing scale anchors are likely to make the respondent think more carefully about the options and hence make a more informed, and most likely more differentiated, choice. Naturally opposing scale anchors are also likely to make the question resemble a classification task, whilst scale anchors referring to level or degree invite the respondent to see the questions as an appraisal or judgement task. The latter is more likely to result in a systematic middle or slightly acquiescent response as respondents are less likely to be able to differentiate their *specific* level of agreement with each individual statement, especially

when presented with a lot of similar questions. As a result we argue that acquiescent and middle response styles can be expected as a reflection of the low level of engagement with the question.

There is some empirical support that scale anchors referring to opposites lead to less acquiescent responses. Some of the items in the Globe study (House, Hanges, Javidan, Dorfman, & Gupta 2004) were constructed in this way, e.g. - ''In this society, people are generally: tough/tender'' or ''I believe that the economic system in this society should be designed to maximize: individual interests/collective interests''. Smith (2004) compared the acquiescent response bias in six large-scale cross-cultural studies. He found that only some of the Globe dimensions correlated highly with other cultural dimensions and hence showed a similar acquiescence bias. We argue that this might well be a scale effect rather than a culture effect. Dimensions that did *not* show significant correlations with other cultural dimensions (e.g. assertiveness and humane orientation) and hence showed no acquiescence bias were measured only with items where scale anchors were opposing words rather than agree/disagree questions. The existing research therefore supports the view that extreme response styles will be stronger when scale anchors refer to naturally opposing and mutually exclusive constructs than when they refer to the level or degree of a construct. In the context of our study's operationalisation this leads to the following hypotheses:

Hypothesis 2a: All five groups will show a significantly lower standard deviation when appraising performance than when classifying typical behaviour as Australian or Chinese.

Hypothesis 2b: All five groups will show a significantly smaller differentiation in appraising performance than in classifying typical behaviour as Australian or Chinese.

Level of knowledge on the topic in question

The level of knowledge (a dispositional factor) an individual possesses about a topic has the po-

tential to influence his/her response style. We anticipate that the less knowledge respondents possess about the topic in question (in our case cultural differences), the less likely they are to commit to a definitive (i.e. more extreme) answer and the more likely they are to stick to the middle of the scale. Hence the likelihood of extreme responses in our study would be expected to be larger for study participants with more knowledge about cultural differences.

There are a number of studies that provide some empirical support for the impact of topic knowledge on response styles. O'Donovan (1965) found that extreme response styles related with the meaningfulness of the questions to the respondent, an effect that was confirmed by Gibbons, Zellner & Rudek (1999). If a respondent has a higher level of knowledge about the subject in question, one would expect the questions to be more meaningful to the respondent. There is also support from the related literature on expertise and extremity. Brauer, Chambres, Niedenthal & Chatard-Pannetier (2004) found that when the task of experts was to identify the best among various choice alternatives – as it is in our case – experts tended to give more extreme evaluations than non-experts.

An important means to increase knowledge about a particular topic is content-specific training, in our case cross-cultural differences in behaviour. In our study, the only group that had received formal training in cross-cultural differences was one of the German groups. This group had followed several courses during their degree that dealt with cross-cultural differences. In addition, the survey was completed in one of their classes in International Management. Hence, in the context of our study's operationalisation this leads to the following hypotheses:

Hypothesis 3a: German respondents **with** cross-cultural training will display a significantly higher standard deviation when classifying behaviour as typically Chinese or Australian than all of the other groups.

Hypothesis 3b: German respondents **with** cross-cultural training will display significantly more differentiation between typical Chinese and Australian behaviour than all of the other groups.

However, cross-cultural training generally also aims to impart an *appreciation* of cultural differences. Therefore, although cross-cultural training might teach students to *identify* cross-cultural differences, it might also teach them to suppress their value judgments about these differences. Therefore, although we do expect the positive knowledge effect on ERS to be present for behaviour, we do not expect it to be present for the performance appraisal questions. Hence:

Hypothesis 3c: German respondents **with** cross-cultural training will not display a significantly lower standard deviation than the other groups in assessing the performance of typical Chinese/Australian behaviours.

Hypothesis 3d: German respondents **with** cross-cultural training will not display a significantly smaller differentiation than the other groups in assessing the performance of typical Chinese and Australian behaviours.

Methods

Sample and data collection procedures

Our data were collected from five groups of students during the first months of 2007. Given that the topic of our study related to typical Chinese and Australian behaviour, the choice of our Australian and Chinese students was straightforward. However, we also wanted to assess whether response styles of Chinese students differed depending on whether they were studying in their own country or abroad, hence we included both Chinese students in China and Chinese students in Australia. Germany is geographically and culturally distant from both China and Australia and hence German students could provide a third perspective. Further, including German students provided opportunity to investigate the impact of knowledge of cross-cultural differences, as our respondents in Germany took an International Business degree.

The use of a student sample poses limitations in terms of representativeness, especially

in developing countries as students might be different from the population as a whole and might be more westernized than non-students. However, this does mean that any cross-country differences in response styles might be attenuated, so in fact our study provides a more stringent test of these differences (Allik & McCrae 2004). A student sample also has some advantages. First, students are a more homogenous group because they are part of the same generation, have the same type of educational experiences and are of a similar age. Second, there are problems with studying response styles with managerial respondents. It is almost impossible to find groups of managerial respondents that are sufficiently closely matched to isolate the phenomenon of interest (response styles). Managerial respondents across countries and organisations are likely to vary on many more dimensions than students and we would be able to control for only a limited number of these dimensions. Hence it would be very difficult to establish to what extent any differences found are due to response styles or simply to variance unaccounted for in the control variables.

Online surveys were used in each group with Anglo-Australian students responding to the questionnaire in English, Chinese students (both in Australia and in China) in Mandarin and German students in German. We built in a filtering question asking respondents about the extent to which they considered their norms and values to be typical of the country they were born in. A response of less than 5 (on a scale of 1-10) on this question automatically triggered the end of the survey and a message to the respondent that unfortunately they were not part of our target audience.

In Australia, data were collected from both Anglo-Australian respondents and Chinese respondents at a large Australian university. Anglo-Australians were defined as those who were born in Australia, the UK or Ireland (the most common Anglo migrants to Australia) and had English as their native language. Any other country/language triggered the end of the survey. Chinese respondents were defined as those who were born in mainland China, Hong Kong, Singapore or Malaysia and had Cantonese or Mandarin as their native language. Any other coun-

try/language triggered the end of the survey. Emails were sent to the postgraduate student mailing list for students enrolled in a professional Masters degree in management. As the questionnaire required quite a bit of time to complete, respondents were given a $25 bookshop voucher as compensation for their effort. Due to resource constraints, we closed the survey after receiving 33 responses for each cultural group. Unfortunately, subsequent analysis showed that two of the Chinese respondents and ten of the Australian respondents had not taken the questionnaire seriously. This was reflected in one of three ways: the same student responding more than once in order to pocket more than one book vouchers, the same answers given virtually everywhere, or a very short completion time. These responses were removed for the final analysis, leaving us with sample sizes of 31 (Chinese students) and 23 (Anglo-Australian students).

In China, responses were collected at a large university in the Jiangsu province. Respondents were MBA students. They were asked in class to complete the questionnaire online in their own time. A reminder email was sent after several days. No incentives were used. Trigger questions similar to those for Australia were built in to ensure we only received responses from Chinese students with Cantonese or Mandarin as their native language and with a score of 5 or more on the typical norms question. We received 36 responses, of which two had to be discarded due to an extremely short completion times.

In Germany, responses were collected at a university in the south of Germany, where students were studying for a Diplom Betriebswirt degree (approximately equivalent to Masters) in International Business. German students often acquire work experience before going to university. In this particular degree, the students typically undertake internships from their first semester of study. We collected data from two groups. The first group consisted of final year students who took the questionnaire in class. Being in their final year meant they had been exposed to several courses that dealt with cross cultural differences. The second group of students were invited in class to complete the online questionnaire in their own time. These were first year students who had not received any systematic instruction in cross cultural differences. For

both groups only minimal incentives were used: two book vouchers in the first group and no homework in the second group. The first group consisted of 24 German students, the second group of 30 students. Neither of the groups showed any evidence of non-serious responses.

Our final sample size is 142 with an average age of 26. Two thirds of the sample was female. Students on average had 3 years and nine months of work experience. The Chinese students in China were on average older and had more work experience, but the other four groups were very similar in their demographics. We will test for the impact of demographic differences between countries in our results section.

Measures

Dependent variables

The dependent variables in our study are extreme versus middle response styles. However, in order to generate data on response styles, we need some content based questions. We focused on typical Australian versus Chinese behaviour, as well as the performance implications of this behaviour. We chose these two countries to portray an Asian and Western culture, because the target respondents could be expected to be familiar with these countries, although the level of familiarity might have differed between our data collection groups. Therefore, we first reviewed the extensive literature on cross-cultural differences for the major areas of difference between Asian and Western cultures. The following three areas of cultural differences were included (text in italics refers to sample items on the survey to test for these cultural differences):

- Emphasis on the group (Asian) versus emphasis on the individual (Western). A sample pair is: *"This employee is humble and will not speak up and share his ideas until his colleagues ask him to"* versus *"This employee will speak up and share his ideas easily, even if that means some people think he is arrogant"*. This area relates to the cultural dimension of Individualism versus Collectivism as identified among others by Hofstede (1980 2001), House et al. (2004) and Trompenaars (1997).

- Respect for authority/loyalty/subservience (Asian) versus independent initiative/challenging the superior (Western). A sample pair is: *"This employee thinks subordinates should give their superiors respect automatically, irrespective of their performance"* versus *"This employee thinks subordinates should give their superiors respect only if they deserve it through their good performance"*. This area relates to the cultural dimension of Power Distance as identified among others by Hofstede (1980 2001), House et al. (2004), and Trompenaars (1997).

- Focus on effort and process (Asian) versus focus on ability and outcome (Western). A sample pair is: *"He will take care to follow the correct procedures, even if this does not always lead to the best outcome"* versus *"He will ignore the correct procedures if this means it will lead to the best outcome"*. There is no clear parallel with existing cultural dimensions although there are some links with Uncertainty avoidance and Masculinity versus Feminity (Hofstede 2001) and Assertiveness (House et al. 2004).

We wrote 15 sets of opposing statements for each of the three cultural dimensions. This resulted in 90 questions consisting of 45 matched pairs that describe opposite behavioural tendencies. The statements were originally constructed in English and then translated to Mandarin by a native speaker and back-translated by another native speaker. Statements that resulted in translation problems were rewritten in English or replaced all together if a satisfactory translation could not be found. The final set of English statements was then used as the basis for the German translation, which followed a similar procedure. After translation, the questions were randomized so that opposing statements did not follow each other and questions with regard to the three underlying cultural differences were interspersed.

In the survey, students were first asked some brief filtering questions (as outlined above) as well as several demographic questions. Thereafter they were presented with the following instruction and 90 statements: *"The next part of this survey consists of a series of short state-*

ments all describing possible behaviours and attitudes of employees working in a company. You will be asked to rank the extent to which you think the statement reflects typical Chinese or Australian behaviours or attitudes, using this scale:"

Australian   1   2   3   4   5   6   7   8   9   10   Chinese

Once the study participants completed the first part of the survey they received the following instruction: *"The final part of this survey again consists of a series of short statements all describing possible behaviours and attitudes of employees working in a company. This time we ask you how you would rate the performance of an employee who displays this type of behaviour or attitude, using this scale."*[1]

Low Performance   1   2   3   4   5   6   7   8   9   10   High Performance

In order to create the dependent variables for our study, we first calculated the standard deviation for all 90 statements dealing with behaviour ("SD behaviour") and all 90 statements dealing with performance ("SD performance"). The higher/lower the standard deviation for a particular respondent, the higher the proportion of his or her extreme/middle responses. This is a better way to assess extreme/middle response styles than to artificially designate part of the scale as extreme or middle responses. Given the length of the scale it would clearly have been inappropriate to only pick the ends and the middle of the scale: any other decision would have involved subjective judgment.[2]

    We subsequently calculated the difference scores for behaviour by subtracting the ratings

---

[1] In retrospect we should probably have mentioned the country context in which respondents were expected to evaluate performance. However, we argue that in the absence of a country specification, respondents would have been most likely to evaluate performance in the context of their home country. This is likely to be true even for the Chinese students in Australia as researchers have found that values change slowly (see Hofstede, 1980/2001 and Olivas-Luján, Harzing & McCoy, 2004).

[2] A reviewer suggested that standard deviation might not be the most appropriate way to measure extreme response styles as – theoretically – a respondent might display a very low standard deviation if he/she picked all of his/her responses at one end of the scale. We therefore reran our analyses reported below with a recoded variable that gave a score of +2 for 1/10 scores, +1 for 2/9 scores, 0 for 3/8, scores, -1 for 4/7 scores and -2 for 5/6 scores. Our results were identical to our analyses based on standard deviation, both in terms of relative ranking of the five data collection groups, and in terms of statistical significances between them.

for "Australian" statements (i.e. those statements that were created to reflect Western/Australian behaviour) from the ratings for "Chinese statements" (i.e. those statements that were created to reflect Asian/Chinese behaviour). The difference scores for performance were similarly created by subtracting the ratings for Australian statements from the ratings for Chinese statements.

Independent and control variables

The independent variables for hypothesis 1 and 3 consisted of the five data collection groups. These groups differed in cultural background (Western versus East Asian) and level of knowledge about the topic of investigation (German group with training versus all other groups). The independent variable for hypothesis 2 was the type of scale anchor: mutually opposing scale anchors (Australian versus Chinese) versus scale anchors reflecting level or degree of a construct (low versus high performance).

Our control variables were age, gender, and time taken to complete the survey. Extreme responses styles have been shown to increase with age (Greenleaf 1992; Ross & Mirowsky 1984) and males have been shown to have a higher level of acquiescence than females (Johnson et al. 1997; Ross & Mirowsky 1984). Baumgarten & Steenkamp (2001) indicate that time pressure can affect response style.

Method of analysis

We used two methods of analysis with the data. First, we generated some descriptive statistics to assess whether we had succeeded in our attempts to create opposing statements of Australian versus Chinese behaviour. Second, we ran a general linear model including both our independent (data collection groups) and control variables, in order to establish whether the control variables had a significant impact on our dependent variables (standard deviation and difference scores). Since this was not the case for any of the control variables, we were able to use a simple Analysis of Variance (ANOVA), with a post hoc Tukey B test, to compare mean scores between

our five data collection groups.

Results

Descriptive results

Table 1 displays the descriptive results of this study. For each of the five groups, we provide the mean score for both the behaviour and the performance questions, as well as the standard deviation for both types of questions. There is no significant difference in mean scores between the five groups, either for the behaviour or for the performance appraisal questions. This is to be expected as the 90 questions consist of 45 matched pairs that describe opposite behavioural tendencies and respondents were asked to score the behaviour as Australian versus Chinese and rate different types of behaviour as low or high performing. One would therefore expect the overall average for each set of questions to be close to the theoretical average of 5.5. As can be easily verified from Table 1, all groups are close to this theoretical average both for the behavioural and for the performance questions.

===============

Table 1 about here

================

Differences between data collection groups in terms of standard deviation, however, are substantial. We will return to these differences when describing the results for our hypothesis testing. Table 1 also includes the difference scores for both behaviour and performance. The positive score for behaviour indicates that for each of the five groups on average those statements that we thought reflected Chinese behaviour (right-hand of the scale) were indeed seen as more reflective of Chinese behaviour than statements that we thought reflected Australian behaviour (left-hand side of the scale).

The consistent negative difference scores for performance show that both Asian and

Western respondents rated Australian behaviour higher than Chinese behaviour. This seems to show a movement towards Western ideals for Asian respondents, even though the perceived typical Chinese behaviour has not yet changed. This can be at least partly explained by the specifics of our sample. Our Chinese students in China were enrolled in an MBA, a Western style management education, usually dominated by Western management theory. Our Chinese students in Australia had chosen to study in a Western country and were taking Western degrees. Figure 1 visually illustrates how Australian and Chinese behaviour are evaluated differently. Both Asian and Western respondents rate Australian behaviour higher than Chinese behaviour, although the difference is much smaller for Asian respondents than for Western respondents.

===============

Figure 1 about here

================

Hypothesis testing

The five samples showed some variability in terms of demographics. The Chinese students in China were on average older, but the other four groups were very similar in their demographics. Therefore, before testing our hypotheses, we ran four general linear models, including a range of demographic control variables as well as the five data collection groups. This would allow us to assess whether control variables should be included in our hypothesis testing. None of the control variables (age, gender, or time taken to complete the questionnaire) had a significant impact on our dependent variables: the standard deviation or mean difference, neither for the behavioural nor for the performance appraisal questions. In fact, none of the demographic variables even came close to reaching significance, with F-values all well below 1.00. Only the data collection group was a significant ($p < 0.000$ in three of the four comparisons, $p < 0.001$ in the fourth) explanatory variable, with F-values ranging from 5.32 to 23.25. Hence we proceeded to test our hypotheses using analysis of variance, which allowed us to compare differences between the five data collection groups.

The first set of hypotheses (1a to 1d) examines differences in response styles between groups of different national backgrounds, contrasting collectivist and individualist respondents and proposing a more compressed response style for collectivist respondents. As is shown in Table 1, East Asian (Chinese respondents in China and Australia) respondents displayed a significantly smaller standard deviation and difference score than Western respondents. This was true for both the typical behaviour and the performance questions. For example, with respect to Hypothesis 1a, we find that the standard deviation for East Asian respondents (1.77 and 2.07) is significantly smaller than the standard deviation for Western respondents (2.52, 2.63 and 3.37). As the different superscripts in Table 1 show, the same is true for Hypotheses 1b-1d. Thus we find full confirmation for hypotheses 1a-1d. In combination, these hypotheses confirmed that collectivist countries demonstrate more compressed responses.

The second set of hypotheses (2a and 2b) assess differences in response styles based on the type of scale anchor used and argued that extreme response styles would be stronger when scale anchors refer to naturally opposing and mutually exclusive constructs than when they refer to the level or degree of a construct. Our associated hypotheses hence predict that all groups will show a lower standard deviation when assessing performance than when classifying typical behaviour. We also predict that all groups will display a lower level of differentiation in assessing performance than in classifying behaviour. Table 1 shows that both predictions are supported for all five groups. For instance, for Chinese students in Australia, the standard deviation for classifying typical behaviour was 2.07, whilst the standard deviation for assessing performance was 1.77. For this same group, the difference score was 1.42 for classifying typical behaviour, whilst for assessing performance it was only 0.24.

A formal test of the significance of the differences in standard deviation between performance and behaviour questions shows that the standard deviation for performance is significantly lower for Chinese in Australia (p = 0.016), for Anglo-Australians (p = 0.021), for Germans without training (p = 0.001) and for Germans with training (p = 0.000). The only group

that does not show a significant difference is Chinese in China, possibly because the standard deviation for behaviour for this group was very low to start with. Therefore, we find substantial support for Hypothesis 2a. A formal test of the significance of the differences in the level of differentiation in assessing performance versus behaviour shows that for all groups this difference is significant at p = 0.000. Hence, we find very strong confirmation for Hypothesis 2b.

Figures 2 and 3 provide a visual illustration of the differences between the five groups in terms of standard deviation and difference score. Figures 2 and 3 also illustrate how both standard deviation and difference scores are systematically lower for performance questions than for behaviour questions. Overall, these results provide considerable support for our second set of hypotheses on the importance of the type of scale anchor to response styles.

===================

Figures 2 and 3 about here

===================

The third set of hypotheses focus on the level of knowledge about a particular topic to the level of extreme responses. We predict that the higher this level of knowledge, the more extreme the responses will be. Specifically, our associated hypotheses predict that the German group **with** cross-cultural training will show a significantly higher standard deviation in their responses about typical Chinese-/Australian behaviour (H3a) and more differentiation between these behaviours (H3b). The German group with training indeed displayed a significantly higher standard deviation (i.e. 3.37) than all other groups (all at p = 0.000, t-values ranging from 5.608 to 11.330). The German group with training also had a significantly higher difference score (i.e. 4.53) than all other groups (all at p = 0.000, t-values ranging from 7.753 to 16.765). We therefore find very strong support for both H3a and H3b.

However, we also argued that cross-cultural knowledge includes not just *knowledge* of cultural differences, but also an *appreciation* of these differences. We predict that the German

group with training will **not** display significant differences in their standard deviation (H3c) and difference score (H3d) for the performance questions. As can be seen in Table 1, German students with training *do* show a slightly higher standard deviation and difference score than the other groups. However, as the identical superscript shows, they are no longer significantly different from the other Western groups. They are still significantly different from the Asian respondents, but this is true for all groups of Western students and hence this is likely to be a cultural rather than a knowledge effect. We therefore also find support for H3c and H3d. Overall, we therefore find strong support for our third set of hypotheses that proposed a positive relationship between the level of knowledge and response styles.

Discussion

In order to test several dispositional and situational determinants on response styles, we asked five groups of respondents (Chinese in China, Chinese in Australia, Anglo-Australians in Australia, and two groups of German students in Germany) to classify and assess typical Australian or Chinese behaviour. We found national/cultural background to have a strong impact with respondents from collectivist countries displaying a more compressed response style. This study also included a new situational variable: the type of scale anchors used. We found extreme response styles to be stronger when scale anchors referred to *naturally opposing and mutually exclusive constructs* than when they referred to the *level or degree of a construct*. Finally, we included a new dispositional variable: the level of knowledge a respondent has about the topic in question. We found that a higher level of knowledge positively influences extreme response styles.

Implications for theory and (cross-national) survey research

Our study has several significant implications for survey research, both in a national and in a cross-cultural setting. It reinforces the conclusion of earlier studies that show that response

styles differ significantly between cultures. As many of these studies were conducted several decades ago, it is important to establish that middle response style is still a very prominent feature of East Asian respondents. Moreover, we also showed that middle response styles are present even when scale anchors refer to mutually opposing constructs, rather than only for scale anchors referring to level or degree as was the case in previous studies. This conclusion is important not just for cross-national comparisons, but also for research within cultures, as many countries are becoming increasingly multi-cultural. Cross-cultural social researchers could easily draw erroneous conclusions, based on between group biases if they do not recognise the impact of response styles. Researchers therefore need to pay more attention to response styles in their data collection procedures as research clearly shows that there is stability in response style differences between countries (see e.g. Harzing (2006) for a recent summary).

It is possible to identify response styles, for example, through multi-group confirmatory factor analysis (Cheung & Rensvold 2000) or approaches based on item response theory (Van de Vijver & Leung 1997). Any response style differences can then be addressed after data collection, e.g. through within person standardization (see, e.g. Leung & Bond 1989; Smith 2004). However, Fischer (2004) argues that this might not always be the best solution and the debate on the advantages and disadvantages of standardisation seems to be ongoing. It would seem much more effective to try to avoid response style problems in the design stage of the research project. As Van de Vijver (2003:233) aptly commented: *"The statistical toolbox of the cross-cultural survey researcher has become both larger and more sophisticated. [..] [However] statistical sophistication in data analysis cannot compensate for poor quality of study design"*. Similarly, Baumgarten & Steenkamp (2001) call for research that would identify response formats that would suffer least from response styles, both within and across cultures. Unfortunately, so far there is very little advice on how to minimise response bias through survey design, but there are some promising developments.

In a cross-cultural comparison of response styles in 16 countries Harzing et al. (2009)

found that changing the number of response options from 5 to 7 provided a slight reduction in both response and language bias. The reduction in response bias confirms earlier results for more limited group of respondents by Hui & Triandis (1989) and Clarke (2001). However, Harzing et al. (2009) found that asking respondents to *rank* rather than *rate* alternatives provided even better results in terms of a reduction of bias. Unfortunately, the statistical tests that can be conducted with ranked data are limited. Chami-Castaldi, Reynolds & Wallace (2008) suggested a very promising alternative: individualised rating scales. In an exploratory study Chami-Castaldi et al. (2008) allowed respondents to define their own rating scales, both conceptually and numerically. Using bespoke online survey software, respondents were provided with a neutral anchor of zero, with its conceptual definition defined to the respondent. Subsequently, they were asked to define endpoints of the scale that represented the respondent's conceptual agreement/disagreement extremes. Conceptual scale anchors chosen included: totally, fully, absolutely, strongly, definitely, and completely. Some respondents chose different scale anchors for the positive and negative extreme, but did express that they reflected similar levels of intensity. Numerical anchors ranged from 2 to 6. This exploratory work clearly established the feasibility of respondents defining their own individualised rating-scales. However, as the authors indicate while it may be *possible* to have individualised rating-scales, the *reliability* and *validity* of this measurement method has not yet been established. Further research in this area would be of great benefit to cross-cultural researchers.

In addition to cross-cultural differences, our results also showed that response styles differ according to two variables that had not been previously investigated. First we showed that extreme response styles were stronger if scale anchors refer to naturally opposing and mutually exclusive constructs than when they refer to the level or degree of a construct. The first type of scale anchors force respondents to more actively engage with the questions and hence their answers are more likely to reflect their true opinion. As the higher level of extreme responding for these type of questions shows, respondents are more likely to provide stronger differentiation

between questions. Although extreme response styles might be a problem if they are artificial or differ substantially by country, they might also be advantageous for researchers if they increase *meaningful* differentiation between respondents. Hence, it would seem appropriate for researchers to look beyond the traditional Likert scales, where scale anchors are based on level or degree. By thinking more creatively about how scale anchors could be constructed to ensure they are mutually exclusive, researchers could make sure that respondents are more likely to engage with the questions and less likely to rely on acquiescent or middle response styles. It is not easy to construct naturally opposing and mutually exclusive scale anchors, especially in a cross-cultural setting. Words that are seen as opposites in some countries might not be opposites in other countries. In addition, some concepts – such as success and money – could be seen as different and unrelated in some countries, but similar and related in other countries. Substantive efforts on the definition of scale anchors will be useful for IB/IM researchers to fully understand the complexities of cross cultural differences and the development of appropriate policy and practice recommendations for multinational organisations.

Second, we found that the level of knowledge respondents have about the topic in question positively influences their ability to distinguish between the two alternatives and hence results in more extreme responses. This result might even explain the East Asian versus Western differences in response styles. It is possible that our questions were simply not meaningful for our Chinese respondents and they did not have a strong opinion on the issues in question. However, this does not explain why *all* studies relating to response styles have found less extreme and more middle responses for East Asian respondents. It is unlikely that all these studies dealt with topics that were not relevant to their East Asian respondents, although it is possible that a Western bias might make questionnaires irrelevant to East Asian respondents. Therefore future researchers should include questions to assess the level of respondent knowledge of a topic. Understanding the level of respondent knowledge will assist in interpreting the results and developing effective policy recommendations.

We will come back to question relevance in the next section. In general though, our findings would lead to the recommendation to ensure that questions are meaningful to respondents, because they have a certain level of knowledge or experience with them. Although this recommendation might sound trivial, a large number of studies expect respondents without substantial life or work experience (e.g. first year undergraduate students) to be able to give meaningful answers to a wide variety of questions. We would therefore strongly caution against the use of student samples, unless studies investigate fundamental processes, structures or outcomes, where we are interested in "*basic characteristics of human nature that are relatively independent of context and life experiences*" (Bello, Leung, Radebaugh, Tung & van Witteloostuijn 2009:362) or discuss topics for which students can be expected to have at least some relevant knowledge, such as every day work behaviours, as we do in this study.

Limitations and suggestions for further research

Our study has several limitations. The first limitation relates to the relatively small size of our samples. Although our results were highly significant, further research with bigger sample sizes would increase confidence in the robustness of our results. The next two limitations are related to our data collection groups. For the German group with training, we cannot be certain whether the stronger differentiation effect was due solely to a greater knowledge of cross-cultural differences. Part of the stronger effect might have been caused by the fact that the survey was completed as in-class task. Hence students might have felt a greater need to be "accurate" in their assessment. Even so, students could only be expected to be accurate if they had sufficient knowledge to accurately classify the different types of behaviour. Hence, we argue that knowledge will at least have played some role, especially given the very large and significant differences for this hypothesis. Even so, future studies might want to design a more strictly controlled comparison between groups with and without extensive knowledge of the subject in question.

We did not find any significant differences between the two groups of Chinese students:

those that studied in China and those that studied in Australia. Even so, the Chinese students in Australia showed a slightly higher standard deviation and difference score for the behaviour questions than the Chinese students in China. In particular, although they had very similar scores for Chinese behaviour, Chinese students in Australia were a bit more likely to classify statements intended as Australian as Australian behaviour. This might have been caused by the fact that these students had been exposed to Australian behaviour and hence had more active experience with (and therefore knowledge of) this behaviour than Chinese students in China. In order to test this hypothesis, it would have been interesting to know how long our Chinese students had been in Australia. Unfortunately, we neglected to include this question in our questionnaire. However, since our target audience were Masters level students, it is most likely that the majority of them had come to Australia for study and hence their length of stay was limited. Again, future studies might want to more systematically study the influence of both experience and knowledge on response styles. If indeed extreme response styles are partly caused by the level of knowledge or experience a respondent has with the topic in question, extreme response styles might be seen as positive. One would normally attach more importance to the opinions of respondents with more knowledge and experience.

A fourth limitation is related to the language of the questionnaires. In the current study, all respondents took the survey in their native language. It would have been interesting to establish whether responding in English for the non-native English speakers would have led to different response styles and whether the level of exposure to English would make a difference in this respect. Although several studies have looked at these and related topics (e.g. Bond & Yang 1989; Gibbons, Zellner & Rudek 1999, Harzing et al. 2005, Harzing et al. 2009), there is still much we do not know about the impact of language on survey responses.

A final limitation is one that our study shares with virtually all other studies in this field: we used a questionnaire designed by a group of Western researchers and applied it in a very different cultural setting. It is possible that the lack of variance in the answers by Chinese students

was simply caused by the fact that questions were not meaningful to them. In order to test this proposition, future studies might consider using a questionnaire originally developed in East Asia and investigate whether differences in MRS and ERS between East Asian and Western countries are smaller when non-Western questionnaires are used. Please note, however, that this "non-indigenous hypothesis" cannot explain why response styles in some other Asian countries such as the Philippines, India, Malaysia and Indonesia (see e.g. Harzing 2006, Stening & Everett 1984) are more extreme than even those of Western respondents. In this respect, Harzing (2006) showed that ethnicity played a more important role than nationality, with Malaysian-born Chinese respondents scoring very different from Malay respondents and similar to respondents from Hong Kong and mainland China.

Conclusion

Conducting research across countries provides an opportunity to investigate the impact of a wide variety of organisational policies and practices. The value of such research, however, can be undermined by systematic differences in response styles. Our study confirms that culture does have an impact on response styles. Respondents with an East Asian (Chinese) background are more likely to use the middle ranges of a scale while respondents with a Western (Australian and German) background are more likely to use the extremes of a response scale.

A new way of defining response anchors and the use of a larger (i.e. ten) scale points resulted in more ERS in our study. We suggest that naturally opposing and mutually exclusive scale anchors encourage respondents to fully engage with questions potentially producing better quality data for the researcher. The implications of this finding are possibly far reaching: many of our existing measures with their reliance on agree/disagree Likert anchors may be poor measures, but the application of new measures undermines our capacity to make comparisons with earlier studies. So we encourage future researchers to conduct replication and extension studies on scale anchors.

Our study demonstrates the impact of knowledge of a topic, particularly cross cultural knowledge. Future researchers should consider including questions in their surveys that assess the level of topic knowledge. Researchers could also usefully investigate the implications of different sources of knowledge. It would be useful to compare the impact of knowledge gained formally through, for example, higher education or workplace training programmes or informally through, for example, travel or personal experiences.

Collecting high quality data is essential for all empirical studies. While response styles pose a challenge to the collection of high quality data, our paper provides researchers with some strategies for managing this challenge.

References

Allik, J., McCrae, R.R. (2004). Toward a Geography of Personality Traits: Patterns of Profiles across 36 Cultures, *Journal of Cross- Cultural Psychology* 35(1), 13–28.

Baumgartner, H. Steenkamp, J.-B. E.M. (2001). Response Styles in Marketing Research: A Cross-national Investigation, *Journal of Marketing Research* 38, 143–56.

Bello, D., Leung, K., Radebaugh, R., Tung, R.L., van Witteloostuijn, A. (2009). From the Editors: Student samples in international business research, *Journal of International Business Studies* 40, 361–364.

Bond, M.H., Yang, K.-S. (1982). Ethnic Affirmation versus Cross-cultural Accommodation. The Variable Impact of Questionnaire Language on Chinese Bilinguals from Hong Kong, *Journal of Cross-Cultural Psychology* 13(2), 169–85.

Brauer, M., Chambres, P., Niedenthal, P.M., Chatard-Pannetier, A. (2004). The Relationship Between Expertise and Evaluative Extremity: The Moderating Role of Experts' Task Characteristics, *Journal of Personality and Social Psychology*. 86(1),5–18.

Chami-Castaldi, E., Reynolds, N., Wallace, J. (2008). Individualised Rating-Scale Procedure: A Means of Reducing Response Style Contamination in Survey Data? *The Electronic Jour-*

nal of Business Research Methods, 6(1), 9–20

Chen, C., Lee, S.-Y. Stevenson, H.W. (1995). Response Style and Cross-cultural Comparisons of Rating Scales among East Asian and North American Students, *Psychological Science* 6(3), 170–5.

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.

Chun, K.-T., Campbell, J.B., Yoo, J.H. (1974). Extreme response style in crosscultural research, *Journal of Cross-Cultural Psychology* 5(4), 465–80.

Clarke, I (2000). Extreme Response Style in Cross-Cultural Research: An Empirical Investigation, *Journal of Social Behavior and Personality,* 15 (1), 137–152.

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

Couch, A., Keniston, K. (1962). Yeas-sayers and nay-sayers: agreeing response set as a personality variable, *Journal of Abnormal and Social Psychology*, 60, 151–174.

Cronbach, L.J. (1946). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10, 3–31.

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.

Gao, G., Kao, K., Ting-Toomey, S. (1998). *Communicating effectively with the Chinese*, London: Sage.

Gibbons, J.L., Zellner, J.A., Rudek, D.J. (1999). Effects of Language and Meaningfulness on the Use of Extreme Response Style by Spanish–English Bilinguals, *Cross-Cultural Research*, 33(4), 369–81

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

328–51.

Hall, E.T. (1976). *Beyond Culture*, New York: Doubleday.

Hamamura, T., Heine, S.J., Paulhus, D.L. (2008). Cultural differences in response styles: The role of dialectical thinking. *Personality and Individual Differences*, 44(4), 932–942.

Harzing, A.W.K. and 32 country collaborators (2005). The Use of English Questionnaires in Cross-National Research: Does Cultural Accommodation Obscure National Differences?, *International Journal of Cross Cultural Management* 5(2), 213–224.

Harzing, A.W.K. (2006). Response styles in cross-national survey research: A 26-country study. International Journal of Cross Cultural Management, 6(2), 243–266.

Harzing, A.W.K. and 26 collaborators (2009). Rating versus ranking: what is the best way to reduce response and language bias in cross-national research?, *International Business Review,* 18(4), 417-432.

Hofstede, G. (1980). *Cultures consequences. International differences in work-related values*, London: SAGE Publications.

Hofstede, G. (2001). *Cultures consequences, comparing values, behaviors Institutions and organizations across nations*, (2nd ed.). Thousand Oaks: Sage Publications.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., Gupta, V. (2004). Culture, leadership, and organizations. *The Globe Study of 62 Societies*, Thousand Oaks: Sage

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

Johnson, T.J., Kulesa, P., Cho, Y.I., Shavitt, S. (2005). The Relation Between Culture and Response Styles. *Journal of Cross-Cultural Psychology*, 36(2), 264–277.

Johnson, T., O'Rourke, D., Chavez, N., Sudman, S., Warnecke, R., Lacey, L., Horm, J. (1997). Social Cognition and Responses to Survey Questions among Culturally Diverse Populations, in L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, D. Trewin (eds) *Survey Measurement and Process Quality*, pp. 87–113. New York: John Wiley and

Son.

Lee, C., Green, R.T. (1991). Cross-cultural Examination of the Fishbein Behavioral Intentions Model, *Journal of International Business Studies* 25(2), 289–305.

Leung, K., Bond, M. H. (1989). On the empirical investigation of dimensions for cross-cultural comparisons. *Journal of Cross-Cultural Psychology*, 20(2), 133–151.

Olivas-Luján, M.R.; Harzing, A.W.K.; McCoy, S. (2004). September 11, 2001: Two quasi-experiments on the influence of threats on cultural values and cosmopolitanism, *International Journal of Cross-Cultural Management*, 4(2), 211–228.

Peng, K., Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, 54,741–754.

Ross, C.E., Mirowsky, J. (1984). Socially desirable Response and Acquiescence in a Cross-cultural Survey of Mental Health, *Journal of Health and Social Behavior* 25(2), 189–97.

Shiomi, K., Loo, R. (1999). Cross-cultural Response Styles on the Kirton Adaptation–Innovation Inventory, *Social Behavior and Personality* 27(4), 413–20.

Si, S. X., Cullen, J.B. (1998). Response Categories and Potential Cultural Bias: Effects of an Explicit Middle Point in Cross-Cultural Surveys, *International Journal of Organizational Analysis*, 6(3), 218–230

Steenkamp, J.B.E.M., Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-national Consumer Research, *Journal of Consumer Research* 25(June), 78–90.

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

Smith, P.B. (2004). Acquiescent Response Bias as an Aspect of Cultural Communication Style, *Journal of Cross-Cultural Psychology* 35(1), 50–61.

Takahashi, K., Ohara, N., Antonucci, T.C., Akiyama, H. (2002). Commonalities and Differences in Close Relationships among the Americans and Japanese: A Comparison by the Individualism/Collectivism Concept, *International Journal of Behavioral Development* 26(5),

453–65.

Trompenaars (1997). *Riding the Waves of Culture. Understanding Cultural Diversity in Business*. London: Nicholas Brealey Publishing.

van de Vijver, F.J.R., Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.

van de Vijver, F.J.R.(2003). Bias and Substantive Analyses, in: Harkness, J.A., van de Vijver, F.J.R., Mohler, P.P. *Cross-cultural Survey Methods*, Hoboken, NJ: John Wiley and Sons, pp. 207–234.

van Herk, H., Poortinga, Y.H.,Verhallen, T.M.M. (2004). Response Styles in Rating Scales: Evidence of Method Bias in Data from 6 EU Countries, *Journal of Cross-Cultural Psychology* 35(3), 346–60.

Watkins, D., Cheung, S. (1995). Culture, gender, and response bias: An analysis of responses to the Self-Description Questionnaire, *Journal of Cross-Cultural Psychology*, 26, 490–504.

Zax, M., Takahashi, S. (1967). Cultural Influences on Response Style: Comparisons of Japanese and American College Students, *Journal of Social Psychology* 71, 3–10.

Table 1: Group differences in assessment of behaviour and performance*

Data collection group	N	Chinese versus Australian behaviour			Performance assessment of Chinese vs. Australian behaviour		
		Mean	*Standard deviation*	*Difference score*	*Mean*	*Standard deviation*	*Difference score*
Asian respondents							
Chinese in China	34	5.42 [a]	1.77 [a]	1.24 [a]	5.55 [a]	1.71 [a]	-0.29 [a]
Chinese in Australia	31	5.27 [a]	2.07 [a]	1.42 [a]	5.40 [a]	1.77 [a]	-0.24 [a]
Western respondents							
Anglo-Australians	23	5.27 [a]	2.52 [b]	2.90 [b]	5.37 [a]	2.15 [b]	-1.13 [b]
German no training	30	5.41 [a]	2.63 [b]	2.48 [b]	5.44 [a]	2.22 [b]	-1.20 [b]
German training	24	5.38 [a]	3.37 [c]	4.53 [c]	5.40 [a]	2.38 [b]	-1.43 [b]

* Scores with the same superscript in each column are not significantly different from each other (post hoc Tukey B test, significance level of 0.05).

Figure 1: Average performance scores for Chinese and Australian behaviour for Asian and Western respondents
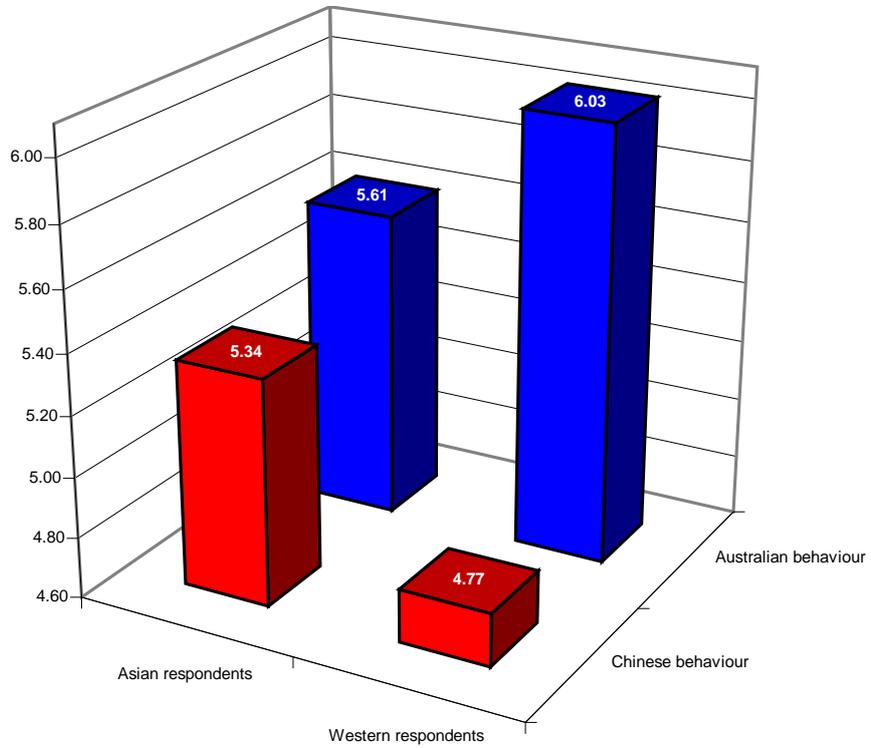
Figure 2: Differences in standard deviation between the five data collection groups for both be-
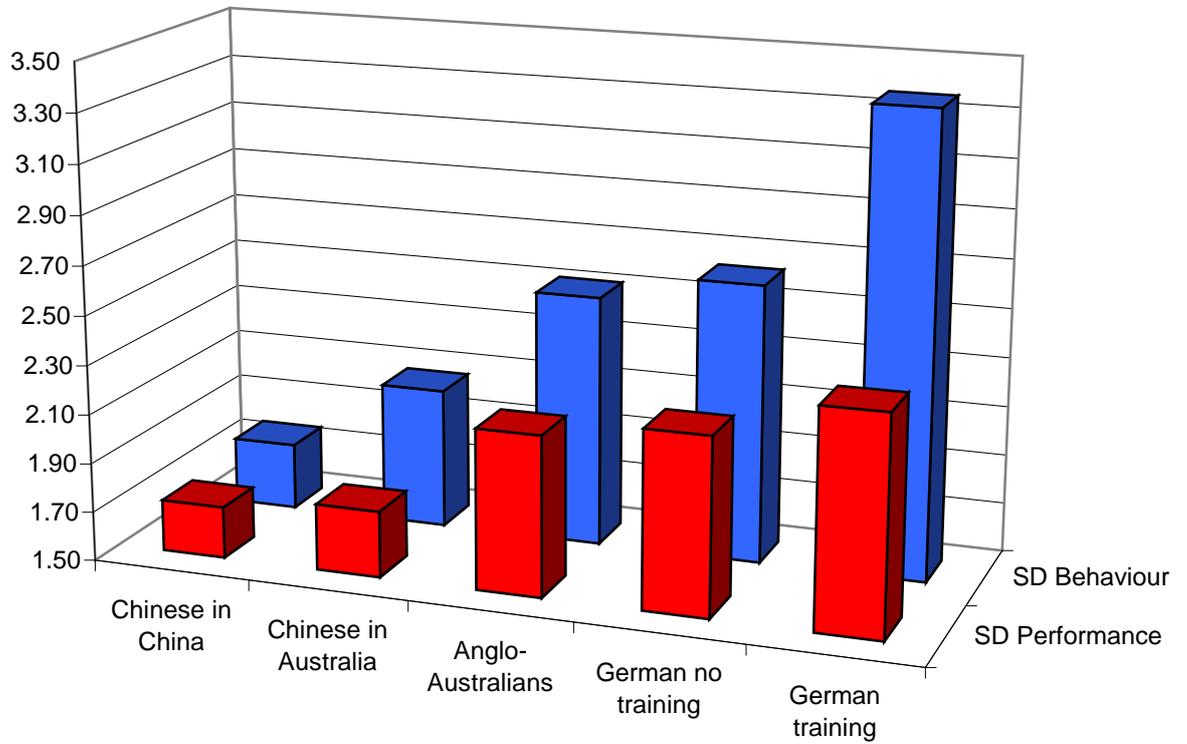
haviour and performance

Figure 3: Differences in difference scores between the five data collection groups for both behaviour and performance