RATING VERSUS RANKING:
WHAT IS THE BEST WAY TO REDUCE RESPONSE AND LANGUAGE BIAS IN CROSS-NATIONAL RESEARCH?

**Anne-Wil Harzing
& 26 co-authors**

Version March 2009

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

RATING VERSUS RANKING: WHAT IS THE BEST WAY TO REDUCE RESPONSE AND LANGUAGE BIAS IN CROSS-NATIONAL RESEARCH?

Anne-Wil HARZING, University of Melbourne, Australia*
Joyce BALDUEZA, Worldbank, Philippines[1]
Wilhelm BARNER-RASMUSSEN, Swedish School of Economics, Finland
Cordula BARZANTNY, Groupe ESC Toulouse Business School, France
Anne CANABAL, University of Maine, USA
Anabella DAVILA, Graduate School of Business Administration and Leadership,
Tecnologico de Monterrey, Mexico
Alvaro ESPEJO, Adolfo Ibañez University's Business School, Chile
Rita FERREIRA, IESE Business School Spain, ESSEC Business School, France
Axele GIROUD, Manchester Business School, UK
Kathrin KOESTER, University of Heilbronn, Germany
Yung-Kuei LIANG, Tatung University, Taiwan
Audra MOCKAITIS, Victoria University of Wellington, New Zealand
Michael J MORLEY, Kemmy Business School, University of Limerick, Ireland
Barbara MYLONI, University of Piraeus, Greece
Joseph O. T. ODUSANYA, HayGroup SA, South Africa
Sharon Leiba O'SULLIVAN, Telfer School of Management, University of Ottawa, Canada
Ananda Kumar PALANIAPPAN, University of Malaya, Malaysia
Paulo PROCHNO, Fundação Dom Cabral, Brazil
Srabani ROY CHOUDHURY, Jawaharlal Nehru University, India
Ayse SAKA, University of Surrey, UK
Sununta SIENGTHAI, Asian Institute of Technology, Thailand
Linda VISWAT, Otemon Gakuin University, Japan
Ayda UZUNCARSILI SOYDAS, Marmara University, Turkey
Lena ZANDER, Stockholm School of Economics, Sweden

---

[1] Co-authors after the first author are listed alphabetically; they all contributed equally to the article. Please note that this list indicates the collaborators' current affiliations, which are not always the universities at which the data were collected. Further some collaborators were not working at that particular university at the time of data collection either, but enlisted local collaborators in their country-of-origin to help with the data collection.

RATING VERSUS RANKING: WHAT IS THE BEST WAY TO REDUCE RESPONSE AND LANGUAGE BIAS IN CROSS-NATIONAL RESEARCH?

ABSTRACT

We propose solutions to two recurring problems in cross-national research: response style differences and language bias. In order to do so, we conduct a methodological comparison of two different response formats – rating and ranking. For rating, we assess the effect of changing the commonly used 5-point Likert scales to 7-point Likert scales. For ranking, we evaluate the validity of presenting respondents with short scenarios for which they need to rank their top 3 solutions. Our results - based on two studies of 1965 undergraduate and 1714 MBA students in 16 different countries - confirm our hypotheses that both solutions reduce response and language bias, but show that ranking generally is a superior solution. These findings allow researchers to have greater confidence in the validity of cross-national differences if these response formats are used, instead of the more traditional 5-point Likert scales. In addition, our findings have several practical implications for multinational corporations, relating to issues such as selection interviews, performance appraisals, and cross-cultural training.

Key words: cross-national research, research methods, response style differences, language bias, survey research.

INTRODUCTION

It has almost become trite to say that the world economy is globalising and that multinational companies play an increasing role on the world scene. However, the resulting increase in interaction between countries has also made both managers and researchers realise that theories and concepts developed in one part of the world (usually the USA) might not be applicable across borders. In order to find out which theories and concepts are universally valid and which have to be adapted, cross-national research is necessary. Oftentimes this type of research is conducted using surveys. However, cross-national survey research is plagued by many problems (for an overview see for instance Singh,

1995; Usunier, 1998; Van de Vijver & Leung, 2000; Tsui, Nifadkar & Ou, 2007). This article focuses on two of these problems: differences in response styles across countries and the possible impact of the language of the questionnaire on the way people respond.

Response styles refer to a respondent's tendency to systematically respond to questionnaire items regardless of item content (Baumgartner & Steenkamp, 2001). Previous research has shown that there might be systematic differences between countries with regard to response styles. This makes a comparison of mean scores across countries a hazardous affair, as conclusions drawn might simply reflect differences in the way people respond to surveys (response style bias) rather than pick up *real* differences in the management phenomena across countries. Although some researchers have suggested that standardisation of responses can be used to remove response styles (see e.g. Leung & Bond, 1989; Smith, 2004), 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.

Research has also shown that the use of English-language questionnaires might create a language bias. Important differences between countries are obscured through a reduced variance in responses between countries, caused either by cultural accommodation or by a lack of the respondent's confidence in responding in a non-native language (see e.g. Harzing et. al. 2005; Harzing, 2006).

Both response styles and language effects are therefore an important threat to the validity of cross-cultural research and it is important to identify methodological solutions to this issue. This article therefore conducts a methodological comparison of two different – but commonly used – response formats – rating and ranking – in terms of their ability to address both response styles and language effects.

Rating normally involves Likert-type scales on which respondents rate their level of agreement or the level of importance of a series of statements on a predefined number of scale points. Although theoretically any number of scale points could be used, in practice most studies use either 5-point or 7-point Likert scales. Ranking can take a variety of forms. In studies of cross-cultural values, ranking has been used quite extensively. For example, studies such as the Rokeach Value survey (Rokeach, 1973) asks respondents to rank 18 values in order of importance. However, this type of ranking has been found to provide unstable and inconsistent results (Peng, Nisbett & Wong, 1997). We suggest that one of the reasons for this may be that respondents are unable to reliably dis-

tinguish among more than a number of different options. Therefore, our study uses ranking in a more structured way. It presents respondents with short scenarios describing a concrete managerial problem with a predefined set of possible solutions and asks them to rank their top 3 solutions. Scenario-based techniques have been shown to provide the best criterion validity (Peng et al., 1997). To date, scenario-based techniques have not yet been assessed in terms of their ability to deal with response styles and language effects.

Therefore, this article employs scenarios to evaluate two specific solutions to the problems caused by cross-country differences in response styles and the possible effect of language on questionnaire response. For rating, we assess the effect of changing the commonly used 5-point Likert scales to 7-point Likert scales. For ranking, we evaluate the validity of presenting respondents with short scenarios for which they need to rank their top 3 solutions. The first involves changing the Likert scale format from the most commonly used 5-point scale to a 7-point scale. The second involves using ranking rather than rating as a response method. Our results, based on two studies conducted with a sample of 1965 undergraduate and 1714 MBA students in 16 countries show that both solutions can reduce response and language bias, but that ranking will generally provide a superior solution in this respect. This finding should allow researchers to have greater confidence in the validity of cross-national differences, if these response formats are used instead of the more traditional 5-point Likert scales. In addition, our findings have several practical implications for management in MNCs, relating to issues such as selection interviews, performance appraisals, and cross-cultural training.

The remainder of this article is structured as follows. In order to substantiate our hypotheses, the next section reviews the literature on response styles and language effects in cross-cultural research. Subsequently, we describe our sample, measures and method of analysis. After presenting the findings of our empirical study, we close with a discussion and conclusion.

LITERATURE REVIEW

RESPONSE STYLES AND 5 VERSUS 7-POINT RATING SCALES

A large number of studies have confirmed that there are substantial and systematic differences in response styles between countries (for recent reviews see Baumgartner & Steenkamp, 2001;

Smith, 2004, Harzing, 2006). The most commonly cited examples of response styles are acquiescence (ARS) or dis-acquiescence (DRS), i.e. the tendency to agree or disagree with an item regardless of the content, and extreme response styles (ERS) versus middle response styles (MRS), i.e. the tendency to use the extreme or middle response categories on ratings scales. Harzing (2006) found these response styles to be related to different cultural dimensions such as power distance, collectivism and uncertainty avoidance (Hofstede, 1980, 2001; House, Hanges, Javidan, Dorfman & Gupta, 2004). Country-level extraversion – as measured by the Eysenck Personality Questionnaire (EPQ) and tapping into aspects such as expressiveness and liveliness – was also shown to impact both positive ERS and ARS.

Of course researchers can always try to remove response bias after the fact by using standardisation (Leung & Bond, 1989; Smith 2004). This procedure has become increasingly popular in cross-cultural studies (Fischer, 2004). However, some of the *true* differences across countries in responses might also be removed in this process. It is very difficult to assess what part of, for instance, a high mean score is caused by an acquiescence bias and what part truly reflects a strong opinion about the subject in question. Fischer (2004) cautions researchers to apply standardisation indiscriminately, especially if they do not have a strong theoretical rationale for why mean differences across constructs should be an indication of bias rather than some meaningful variation. Furthermore, standardisation is only recommended when the questionnaire measures a large number of constructs. If a questionnaire includes only a small number of constructs, the respondent's response style might reflect responses to a meaningful construct rather than a response bias (Hanges & Dickson, 2004).

This paper therefore looks at this thorny issue from a different angle and suggests two alternatives to mitigate or eliminate response styles during data collection, rather than remove them after the fact. One solution that has been proposed - in both the Marketing and Psychology literature - to mediate the impact of ERS in particular is to use Likert scales with a larger number of categories. This allows respondents with a relatively strong opinion to voice a more nuanced position, rather than being forced to choose the most extreme answer. At the same time, the extreme responses that are given on a scale with a wider range of options become more meaningful. However, although the advantage of 7-point scales over 5-point scales in this respect is well-established, it is still unclear whether the same advantages apply to a non-US setting. Very few studies so far have looked at this

issue in a cross-cultural setting. Two rare exceptions are Hui and Triandis (1989) who found that ERS for Hispanics disappeared when 10-point Likert scales were used and Clarke (2001) who found ERS to decline with a larger number of categories for all four countries included in his study (US, Mexico, Australia, France).

In addition to the effect that increasing the number of answer alternatives might have on ERS, we argue that expanding the Likert scale might also reduce the occurrence of MRS. If respondents are able to voice a rather mild level of agreement or disagreement (as portrayed by 3 or 5 on a 7-point Likert scale) rather than being forced to voice the relatively strong level of agreement or disagreement represented by the 2 or 4 on a 5-point Likert scale, they might be less likely to resort to the neutral middle response. Hence:

Hypothesis 1a: Responses to 7-point Likert scales will include a lower proportion of middle responses and extreme responses than responses to 5-point Likert scales.

The provision of additional answer alternatives through 7-point Likert scales provides respondents from countries with a higher incidence of ERS (e.g. Latin American countries, see Hui & Triandis, 1989; Marin, Gamba & Marin, 1992; Harzing, 2006) with an option to express a relatively strong (dis)agreement without having to resort to the scale extremes. It also provides respondents from countries with a higher incidence of MRS (e.g. East Asian countries, see Chen, Lee & Stevenson, 1995; Takahashi, Ohara, Antonucci & Akiyama, 2002; Harzing, 2006) with an option to express a relatively mild (dis)agreement without necessarily having to resort to the middle of the scale. We would therefore expect that differences *between* countries for both ERS and MRS would be smaller when a larger number of answer alternatives are provided.

Hypothesis 1b: Differences between countries with regard to the proportion of middle responses and extreme responses will be smaller for 7-point Likert scales than for 5-point Likert scales.

RESPONSE STYLES AND RANKING VERSUS RATING

Whereas increasing the number of answer alternatives might alleviate the incidence of MRS and ERS, it does nothing to reduce ARS or DRS, since these response styles are simply defined as the proportion of positive and negative responses. If MRS is reduced, ARS and DRS might even increase. Moreover, appropriate translation of Likert scale anchors into other languages is often very

difficult. Even if they do translate into appropriate local equivalents, the intensity associated with these equivalents may be different from the original language.[2] In our study, we therefore set out to explore several alternatives for the traditional Likert scales.

Instead of having scale anchors reflect levels of importance or (dis)agreement, they can also be incorporated into the question and reflect opposites. Some of the items that were used 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". This would make the "right answer" less obvious and hence would be likely to reduce acquiescent response bias. It would also force respondents to consider each question very carefully as most scale anchors would be different, possibly resulting in answers that are more reflective of the respondent's true opinion than response styles. The disadvantage of this option is that a respondent's interpretation of the questions would often be framed by single words, which is problematic since 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.

It is also important to note that – in spite of a differential scale format – the Globe study did not manage to avoid response styles completely. 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) were measured *only* with items where scale anchors were opposing words rather than agree/disagree questions. Dimensions that correlated highly with other cross-cultural studies (e.g. family collectivism and uncertainty avoidance) were those that were at least partially measured with a "normal" agree/disagree Likert scale. Family collectivism "should be" for instance showed strong and signifi-

---

[2] A particularly striking example is provided by Voss, Stem, Johnson and Arce (1996) who show that while the magnitude estimates for good and very good were 74 and 87 in English, they were 91 and 101 in the equivalent Japanese translation.

cant correlations with two of the five other studies; family collectivism "as is" showed very strong and significant correlations with **all** of the five other studies. Family collectivism "should be" was measured with two agree/disagree items and two items referring to importance or level, family collectivism "as is" was measured with four agree/disagree items.[3] Hence the type of scale anchors used might be a more relevant determinant of the correlations found between studies than the cultural dimensions in question. Scales with agree/disagree and importance anchors show a significant acquiescence bias *regardless* of the study in question.

In this study, we therefore used another remedy, namely to ask respondents to *rank* statements rather than use Likert scales. This technique has been used in some studies comparing cultural values across countries (e.g. Lenartowicz & Roth, 2001) and studies using Rokeach value statements. Ranking generally requires a higher level of attention than rating, as all answer alternatives have to be considered before making a choice. As a result, ranking might lead to higher data quality (Alwin & Krosnick, 1985). However, asking respondents to rank more than a handful of statements puts a very high demand on their cognitive abilities, and might lead them to discard the questionnaire altogether. In this study, we therefore constructed short scenarios with a range of proposed answer alternatives, and asked respondents to select what they thought were the *three* best answer alternatives. This by definition removes response styles such as ERS, MRS, ARS and DRS. Therefore, we argue that using a comparison based on scenarios with ranked solutions will be more likely to accurately depict differences across countries than comparisons based on ratings of statements using Likert scales.

In order to test this, we will compare the clustering of the countries based on rating with the clustering based on ranking. The Globe study includes a very detailed discussion and integration of previous country cluster research (Gupta & Hanges, 2004), including Ronen and Shenkar's (1985) seminal article that provided the earliest reference to country clusters. Therefore, we will compare our clusters (created by both our rating and ranking approaches) with the clusters defined in this Globe study. Of the ten culture clusters defined in the Globe study, the following nine have country

---

[3] Please note that in six of the nine Globe cultural dimensions, one or two (out of four or five) items differed between the values and practices scale of the same dimension. In the case of family collectivism, two out of the four items were different.

representation in our study: Anglo (UK, Ireland, US, Canada), Latin European (Portugal), Nordic (Sweden, Finland), Germanic (Germany, Netherlands), Eastern European (Lithuania, Greece), Latin America (Brazil, Chile, Mexico), Middle Eastern (Turkey), Southern Asia (India, Malaysia, Philippines, Thailand), and Confucian Asian (Taiwan).

Hypothesis 2:   *Country clusters based on ranked responses will conform more closely to the Globe country clusters than country clusters based on Likert scale responses.*

THE IMPACT OF LANGUAGE ON RESPONSE STYLES

A second problem that has been identified in doing cross-national research is the possibility that the language of the questionnaire might influence the way people respond. Recent research has found that when English-language questionnaires were used, there was less variance between countries than when questionnaires in the native language were used (Harzing et al. 2005). Hence our first language related hypothesis will test whether this finding is replicated in our study:

Hypothesis 3a: Differences between countries will be larger for native-language questionnaires than for English-language questionnaires.

Some studies have identified cultural accommodation as a reason for response effects associated with different languages (see e.g. Bond & Yang, 1982; Harzing et al. 2005; Ralston, Cunniff & Gustafson, 1995). However, Harzing (2006) suggested that the reduced variance for English-language questionnaires might also be due to an increased tendency to "sit-on-the-fence" in a foreign language, because the respondent lacks the linguistic confidence to give a decisive answer. Respondents' lower confidence in the foreign language might lead to a preference for more neutral responses, whereas their higher level of confidence in their native language might lead to more extreme responses. The finding that English language competence was positively (negatively) related to ERS (MRS) (Harzing, 2006) supports this assumption. We expect that, when comparing English-language questionnaires with native-language questionnaires, we would see a smaller reduction in differences between countries when using a 7-point scale than when using a 5-point scale. We expect this result because a 7-point Likert scale provides a wider range of relatively neutral and relatively extreme responses (which reflects essentially the same logic for ERS/MRS-avoidance that we outlined in relation to Hypothesis 1a):

9

Hypothesis 3b: The reduction of between-country differences when comparing English-language questionnaires with native-language questionnaires will be smaller for 7-point Likert scales than for 5-points Likert scales.

In the case of ranking, MRS and ERS effects are not present. However, it is possible that a lower level of understanding of the foreign language might lead to a less consistent and more random response, hence reducing variance between countries. On the other hand, comprehension in our scenarios would be likely to be enhanced by the context provided in the scenario and the fact that the solution statements are generally longer than statements used in Likert scales. The provision of context is one of the key reasons Peng et al. (1997) suggested scenarios as a superior solution in terms of criterion validity. We build on this result by investigating whether context also helps to avoid a language effect. Hence, we propose:

Hypothesis 3c: The reduction of between-country differences when comparing English-language questionnaires with native-language questionnaires will be smaller for questions using ranking than for questions using Likert scales.

METHODS

SAMPLE AND DATA COLLECTION PROCEDURES

Data for the questions using 5-point Likert scales were collected in a project conducted between 2001 and 2003. Data for the questions using 7-point Likert scales and the ranking of scenario solutions were collected in a project conducted between 2005 and 2006. This means that the two samples were separated in time by 2-5 years. However, the cultural values that this study taps into have been argued to change only very slowly (see e.g. Hofstede, 1980, 2001; Schwarz, 1999) and hence we consider it very unlikely that this short passage of time would compromise our comparisons.

For both projects, the project coordinator recruited country collaborators through personal contacts and networking at professional conferences such as the Academy of Management. All country collaborators received a 15-page document containing very detailed instructions about the aim of the study; items and constructs; results of the pilot study; translation, data collection and data entry procedures; as well as agreements about co-authorship. All collaborators received access to the final

data set. A document with personal introductions of all collaborators was prepared to promote group cohesion and facilitate networking among collaborators. There was very frequent contact between the project's coordinator and the different country collaborators with any issues and preliminary findings immediately distributed by email to all collaborators. In the second project, for example, more than 2,000 emails were exchanged to ensure harmony and coherence throughout the research team.

In the second project, some countries were dropped and others added. Hence only the 16 countries that were covered in both studies (Brazil, Chile, Finland, Germany, Greece, India, Lithuania, Malaysia, The Netherlands, Mexico, Portugal, Sweden, Taiwan, Turkey, United Kingdom and the United States) were included in the tests that directly compared the two studies.[4] In the test of Hypothesis 2, which only compares results from the second project, four additional countries (Canada, Ireland, Philippines and Thailand) that were only covered in the second study were included.

In the first study, respondents were final year university students following a course in Business Administration, Business & Management, Commerce or a similar subject. They were generally between 21 and 22 years old. The gender distribution varied from 27% female in India to 70% female in Lithuania, with an average of 49% female responses. International students were excluded from our sample, so that our comparisons only included students who could be assumed to be representative of the country in which they studied. Although data were collected on a voluntary basis, response rates were high, generally between 80-100%. Inducements such as class credit or sweets were used in some countries to encourage participation. The resulting sample sizes ranged from 95 for Chile to 147 for Portugal, but for most countries were around 100. Data were collected in-class with a paper and pencil survey between March 2001 and April 2003.

In the second study, respondents were MBA students. Their age varied between 25 years for India and 39 years for the Netherlands, with an average of 32 years old. Work experience similarly varied from just over two years for India to 16 years for the Netherlands, with an average of 9 years. Gender distribution ranged from 8% female in Chile to 73% female in Lithuania, with an average of

---

[4] Data were also collected in France and Japan. However, for a variety of reasons, samples in these countries were not comparable to the other countries in terms of demographics and data collection procedures. Hence these two countries were removed from further analysis.

37% female respondents. Again, international students were excluded from our sample and response rates were in the order of 80-100%. Inducements such as class credit or sweets were used in some countries to encourage participation. The resulting sample sizes ranged from 41 for the Philippines to 168 for Portugal, but for most countries were around 100. Data were collected with a paper and pencil survey in-class between September 2005 and May 2006. Feedback from students in both studies was very positive; many students indicated in their comments that they found the survey interesting and thought it was well written and/or translated.

The use of student samples in both studies poses limitations in terms of representativeness. Especially in developing countries, 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 might be attenuated, so that in fact our study provides a more stringent test of these differences (Alik & McCrae, 2004).

MEASURES

Two types of questions were used for this article. In study 1, students were asked to assess on a 5-point Likert scale the level of importance of various work values in their ideal job after graduation, e.g. "have an opportunity for high earnings", "be consulted by your direct superior in his/her decisions", "have friendly colleagues who help each other", "have security of employment". These questions were adapted from Sirota & Greenwood (1971) and Hofstede (1980) and responses have been shown to differ substantially across countries. A total of 18 questions were included in the questionnaire. In study 2, the Likert scale was expanded to 7 points and the number of questions was expanded to 29, adding questions such as "have a job that is close to where you live", "have a lot of autonomy in your job", "be your own boss", "do something you are really passionate about".

The second study also included six short scenarios that dealt with various aspects of management such as employee reward strategy, decision-making style, the role of the manager, relationship with superior, conflict management style and attitude to employee problems. The scenarios and their solutions, as well as the additional work values questions were developed in three rounds of focus groups each including 6 MBA or PhD students of different nationalities. The management scenarios were preceded by a warm-up scenario that asked students how they would normally address their lecturer. In this article, we only use four of the scenarios to ensure that the number of statements

(4 times 7) to be compared was similar for the work values questions and scenarios. The scenarios dealing with decision-making style and conflict management style were excluded as we felt that the answer alternatives for these scenarios were not as clearly differentiated as those in the other scenarios. In order to be able to conduct tests comparing mean scores, the responses for the scenarios were recoded so that the answer ranked 1 received a weight of 3, the answer ranked 2 a weight of 2 and the answer ranked 3 a weight of 1; all other answers received a weight of 0.

Middle response style (MRS) was calculated as the proportion of questions that received a middle (3 on a 5-point scale or 4 on a 7-point scale) response for each respondent. Similarly, extreme response style (ERS) was calculated as the proportion of responses at the end of the scales (1 and 5 or 7). The level of acquiescent response style (ARS) was calculated by dividing the number of questions that received a 4-5 or 5-7 response by the total number of questions for each respondent. Disacquiescent response style (DRS) was calculated in a similar way, using the number of questions that received a 1-2 or 1-3 response. As we are interested response style patterns, not in the scoring on individual questions or constructs, we did not construct scales. Response styles were calculated using all questions, so that each item had an equal contribution to the composite response style variables.

TRANSLATION

The procedures for translation differed slightly between the two studies. In the first study, the bilingual country collaborators were responsible for the translation of the original English questionnaire. Translations were conducted using translation-back-translation procedures. The translator and back-translator were separate individuals who did not enter into a discussion until after they had finished their translations. Discussions between translator and back-translator usually resulted in the change of some of the translations. Where difficulties remained, a third bilingual person was consulted. The back-translated versions were verified by the project coordinator for consistency across languages, which usually resulted in further changes and discussions between translator and back-translator. For several of the European languages, the project coordinator provided independent verification of the translated versions.

In the second study, the translations were conducted by bilingual research assistants under the supervision of the project coordinator. The translated version was subsequently discussed in a focus group including both the translator and two or three other bilingual students in the presence of the

project coordinator. The other students were instructed to read the translated instrument sentence by sentence and indicate whether the text sounded "natural" to them. Subsequently, they were instructed to look at the original English sentence and assess its equivalence to the native version. If the sentences were not felt to be fully equivalent a better translation was sought through discussion among the participants. Where necessary, the project coordinator provided feedback on the meaning behind the questions. This process took at least three hours, but for some languages (e.g. Japanese, Chinese) it took several sessions lasting up to eight hours in total. In addition, country collaborators verified the surveys for accuracy of translation and for potential local language differences, e.g. Canadian French, different variants of Spanish used in the different Latin American countries.

As in any multi-country study, it is very difficult to guarantee translation accuracy with absolute certainty. However, as a result of the effort invested in the translation process and in view of the very positive feedback about the questionnaire by many of our respondents, we are quite confident that the resulting questionnaires are equivalent in meaning across languages. Further, any potential remaining translation inaccuracies would be attenuated by two factors. First, we will be looking at aggregated response patterns for a total of 18-29 items or scenario solutions and hence translation inaccuracies in *one* item would not have a major impact on overall results. Second, we are looking at the between country differences across 16-20 countries, so that any translation inaccuracies for specific languages would not have a major impact on overall results.

Questionnaires were completed in either English or the native language of the country in question. Collaborators were instructed to make sure that the different language versions were randomly distributed. In most countries, English and native language questionnaires were distributed in the same class. In the remaining countries, different classes of the same module or a related module were used to separate English and native language questionnaires. Respondents were not allowed to choose which language version they completed. An equal number of English-language and native-language questionnaires were distributed. To verify whether collaborators had succeeded in the randomisation process, we tested whether the two language groups differed on the question: *"How similar are your norms and values to the majority of people in your birth country?"* None of the 24 countries in the 1st study or the 18 countries in the 2nd study showed a significant difference between the language versions on this question. Hence, we can be reasonably confident that the two

groups in each country were equally typical of their home country – which was important as our questions dealt with values and behaviours that have been shown to differ across cultures – and only differed from each other in terms of the language of the questionnaire.

ANALYSIS

One-way ANOVA was used to compare countries on response styles and mean responses to the Likert-scale work values questions. Its non-parametric equivalent, the Kruskal-Wallis H test was used to compare countries on their mean scores for the different scenario solutions. Hierarchical cluster analysis was used to test Hypothesis 2, assessing how countries clustered together. Data for the work values questions and scenario solutions were first aggregated by country and were subsequently subjected to a hierarchical cluster analysis using Ward's method, which is designed to optimize the minimum variance within clusters. T-tests were used to compare MRS, ERS and standard deviation between the different language versions.

RESULTS

RESPONSE STYLES AND 5 VERSUS 7-POINT RATING SCALES

In order to assess whether the change to a 7-point scale had an impact on the various response styles (Hypothesis 1a) we compared the four response styles on a country-by-country level between the two surveys. We only included questionnaires completed in the native language in each country, so that we could separate the ranking/rating effect from the language effect. As expected all countries showed a reduction in MRS and overall the proportion of middle responses changed from 21.2% to 13.2% (see Table 1). Whereas the first is above the expected proportion (20%, 100%/5pt scale), the second is below the expected proportion (14.3%, 100%/7pt scale). Overall ERS declined from 28.2% to 23.2%. Hence, we find confirmation for Hypothesis 1a. The picture for ARS and DRS was more mixed. As expected the overall ARS and DRS increased, since they now make up a larger part of the scale (4-7 instead of 4-5 and 1-3 instead of 1-2). However the balance between ARS and DRS, which by some authors is seen as the best measure of acquiescence bias remained very similar.

Table 1: Differences in response style between 5-point and 7-point Likert scales

Type of questions	MRS	ERS	ARS	DRS	ARS-DRS	F-value MRS country differences	F-value ERS country differences	F-value ARS country differences	F-value DRS country differences
Work values 2002 5-point scale	21.2%	28.2%	67.9%	10.3%	57.6%	6.758 ***	11.013 ***	15.435 ***	21.364 ***
Work values 2006 7-point scale	13.2%	23.2%	71.5%	14.6%	56.9%	2.090 **	6.258 ***	11.080 ***	14.012 ***
Proportionate decline/increase	-38%	-18%	+5%	+42%	-1%	-69%	-57%	-28%	-34%

** $p < 0.01$, *** $p < 0.001$

Hypothesis 1b proposed that differences *between* countries with regard to the proportion of middle responses and extreme responses would be smaller for 7-point Likert scales than for 5-point Likert scales. As the F-values in Table 1 show, the differences *between* countries for both MRS and ERS were indeed significantly smaller when a 7-point scale was used. This confirms Hypothesis 1b[5]. Country differences in ARS and DRS also declined when 7-point scales were used, but proportionally less so than differences in MRS and ERS.

RESPONSE STYLES AND RANKING VERSUS RATING

To test Hypothesis 2, we conducted a hierarchical cluster analysis for both the scenario data and the work values data in the second study. In order to separate out the ranking/rating effect from the language effect, we only included questionnaires completed in the native language in each country. Figure 1 produces the resulting dendrogram for the scenario data, whereas the dendrogram for the work values data is produced in Figure 2. In Figure 1, our countries cluster very distinctly. Moving from the right to the left the first split is in between Western versus Asian and Central/Eastern European countries. The second split divides the "English Asian" countries – countries that have English as (one of) their official language(s) – joined by Lithuania, from a more heterogenous mix of countries:

---

[5] Since in both studies the age and gender distribution was quite dissimilar across the country samples, we conducted a multivariate analysis to determine the impact of these demographic variables in comparison to the country of data collection as an explanatory variable. Whilst country of data collection was a very significant explanatory variable for all four responses styles, age did not show a significant relationship with response styles for either of the two matched samples of sixteen countries. However, gender did show some impact on extreme response styles in both samples. We therefore repeated the analysis reported in Table 1 for male students (the biggest group) only, but the results were very similar.

Taiwan, Thailand, Turkey and Greece. In the top half of the dendrogram, the third split divides the Latin American countries, joined by Portugal, from the Western European and North American countries. The four Anglo countries form a tight cluster.

Comparing our clustering with the clusters in the Globe study reveals considerable similarity. The Anglo cluster is clearly present, comprised of Canada, the USA, Ireland and the UK. The Nordic and Germanic cluster are not clearly separated, but these clusters showed little differentiation in the Globe studies as well and in many studies (e.g., Hofstede, 1980, 2001) the Netherlands takes up a middle position between Germanic and Nordic countries. The Latin American cluster is clearly distinguishable, although it is joined by Portugal that is closely linked to its former colony, Brazil. Given that Portugal is the only Latin European country in our study, it is not surprising that it grouped with the culturally similar Latin American cluster. The clustering in the bottom of the dendrogram is less distinct, which is probably partly due to the fact that we only had one or two countries representing most of the relevant Globe clusters (Confucian, Middle Eastern, Eastern European). However, we still find most of the South East Asian countries to cluster together. Finally, given their shared history, it should come as no surprise that Greece and Turkey show considerable similarity.

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

Figure 1 and Table 2 about here

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

The accompanying proximity matrix (Table 2) quantifies the distance between the individual countries on a scale from 0 to 1 and again confirms established country clusters. Brazil's closest neighbours are Portugal and Mexico, the Netherlands confirms its cross-roads position between Germanic (Germany), Nordic (Sweden) and Anglo cultures (USA, Canada); the USA is closest to Ireland and Canada; the "English Asian" countries India, Malaysia and the Philippines are quite similar to each other. In contrast to earlier studies, there is a rather small difference between the USA and the Latin American countries. This is likely to be partly due to the fact that data in the USA were collected in El Paso, located on the Mexican border. However, other Anglo countries (Canada, Ireland and the UK) also show a relatively small distance to the Latin American countries. It is possible that these cultural clusters have grown closer together in the past decades with the economic devel-

opment of the Latin American countries and increased economic interaction through agreements such as NAFTA.

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

Figure 2 and Table 3 about here

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

The dendrogram produced based on the work values data (Figure 2) shows far less differentiation. Moreover, countries that in other studies (including our own scenario data) have been clearly identified as being similar – such as the four Anglo countries and the three Latin American countries & Portugal – are spread over very different clusters. We argue that this is likely to be caused by the fact that the underlying rationale for the clustering seems to be similarity in *response styles*, rather than similarity in *cultures*. The accompanying proximity matrix (Table 3) reinforces this observation. A strong case in point is the USA. Whereas for the ranking data, its closest neighbours were Ireland, Canada and the UK, for the rating data its closest neighbours are Malaysia, Turkey and Thailand. Harzing (2006) already showed that the USA had an acquiescence balance (ARS – DRS) which was much higher than that of Northern and Western European countries and similar in magnitude to that of the Southern European countries.

Overall, we therefore find considerable support for Hypothesis 2: country clusters conform more closely to the clusters identified in the Globe study when using ranking than when using rating.

IMPACT OF LANGUAGE

Following Harzing et al. (2005) we hypothesised that differences *between* countries would be larger for the native-language version of the questionnaire items than for the English-language version. In order to test this we ran a split-sample (split by language) ANOVA analysis for the 2002 and 2006 work values questions and a split-sample non-parametric K-independent samples test of variance for the scenario solutions. So in each case we compared the differences *between* countries for the native language version with the differences *between* countries for the English language version. As Table 4 shows, our results confirmed Hypothesis 3a. In *each* of the three comparisons the overall average F-value or Chi-square statistic was higher in the native version compared to the English-language version, hence indicating that differences between countries are larger when native-language questionnaires are used.

After establishing this base line result, we subsequently looked at the reduction in differences between countries in more detail. In Hypothesis 3b we argued that using 5-point Likert scales would lead to a stronger reduction in differences *between* countries for the English-language version than using 7-point Likert scales. If confirmed this would mean that 7-point Likert scales are more "resistant" to the language effect, hence presenting a preferable choice for international research conducted with English-language questionnaires. We find only partial confirmation for Hypothesis 3b. Confirmation of this hypothesis can be found in the fact that the average reduction in F-value is lower for 7-point Likert scales than for 5-point Likert scales. However, the proportion of items with a reduction in F-value significant at 0.01 is similar. The items that show the largest reduction in variance are similar too: "Be challenged by your work", "Work according to clear and fixed rules and procedures", and "Be able to serve your country".

Table 4: Between-country variance for native-language and English-language versions

Type of questions	F-value/Chi-square value native language	F-value/Chi-square value English	% reduction in F-value-/Chi-square value	% of items with reduction significant at $p < 0.01$
Work values 2002 5-point scale	13.31***	7.24***	46%	72%
Work values 2006 7-point scale	10.14***	6.94***	32%	69%
Scenarios 2006	60.63***	42.11***	31%	25%

*** $p < 0.001$

Finally, in Hypothesis 3c we expected ratings to show a stronger reduction in differences *between* countries for the English-language version than rankings. If confirmed this would mean that rankings are more "resistant" to the language effect than ratings. As Table 4 shows this is largely borne out in our results. Although the reduction in Chi-square value for the scenario rankings is similar to the reduction in F-value for the 7-point Likert scales (though smaller than for 5-point scales), the proportion of items that has a reduction significant at 0.01 is *substantially* lower for rankings. Hence, we find confirmation for Hypothesis 3c and rankings can indeed be expected to be more robust than ratings in dealing with the language effect in international surveys.

DISCUSSION AND CONCLUSION

Our results showed that even though changing from a 5-point Likert scale to a 7-point Likert scale reduced MRS and ERS. This confirms the earlier studies by Hui and Triandis (1989) and Clarke (2001) with regard to ERS for a much larger and varied group of countries. However, it did not fully eliminate the problem of differences between countries in their tendency to use middle or end points of the scale. As expected, switching to a 7-point scale did very little to address the major differences between countries in terms of DRS and ARS. Hence the risk of attributing substantive country differences to what might simply be differences in response styles still looms large. Of course, researchers can always try to remove response bias after the fact by standardisation, but as we have discussed above, this is by no means an uncontested option. Our results suggest that ranking provides an excellent alternative, as it completely eliminates both MRS/ERS and ARS/DRS. Our hierarchical cluster analysis showed that whereas country clusters that were formed based on the ranked scenario solutions provided meaningful results, clustering based on the rated work values questions were in contrast with previous research and appeared to mainly reflect differences in response styles.

With regard to language effects, we hypothesised that differences between countries would be larger for the native-language version of the questionnaire items than for the English-language versions. Our results confirmed this hypothesis and hence the earlier study by Harzing et al. (2005). In each of the three instances, the overall average F-value or Chi-square statistic was higher in the native-language version compared to the English-language version, but the reduction in variance was smaller for 7-point Likert scales than it was for 5-point Likert scales. However, we found that ranking showed an even better performance with regard to language effects. The reduction in variance for the English-language version was very small indeed and only a quarter of the solutions showed a significant reduction in variance.

IMPLICATIONS FOR RESEARCH AND PRACTICE

Our results have considerable implications for cross-national research. We showed that response style and language effects can be attenuated by the use of Likert scales with a larger number of answer alternatives and even more so by the use of rankings. This finding allows researchers to have

greater confidence in the validity of cross-national differences if these response alternatives are used instead of the more traditional 5-point Likert scale. As a result, researchers can focus on studying cross-national differences in substantive issues, rather than being hampered by differences caused simply by the response format of the instrument used in the study. We also encourage researchers to revisit previous inconclusive findings (or small significance and/or effect sizes), since they might be able to better measure their constructs by using rankings instead of the typical 5-point Likert scales.

This study also holds a number of practical implications for situations that require managers in multinational corporations to employ rating procedures. Structured selection interviews are one example. Knowing that 7-point rating scales are more culturally reliable than 5-point rating scales may be helpful for enhancing the consistency of recruiters' selection decisions within multinational companies. In particular, one could imagine situations where situational interviews (e.g., Latham, Saari, Pursell & Campion, 1980) might be cost-effectively "automated" by having respondents select their preferred behavioural response. The ranking procedures described here (particularly the most-least procedure described below) are geared to assess the respondent's preference hierarchy among a set of discrete (behavioural) options. Thus, it would be interesting for future research to explore whether automating situational interviews (using a regular ranking or most/least ranking response approach) would provide a more thorough and informative assessment (of candidates' judgment and organisational fit) than would the traditional situational interview approach that merely solicits the candidate's top behavioural preference.

Another context in which the ranking approach may be more useful than the traditional likert-scale approach is that of cultural assimilator training (e.g., Brislin; 1986; Cushner, 1989). Culture assimilator training requires trainees to select a preferred behavioural response, receive feedback on that response, and then (if the initial response was inappropriate) select a different behavioural response to the same situational question until the trainee grasps the cultural principles underlying the best response. Although there is usually only one best response to such questions, and immediate feedback per response is likely to be most effective during the training process itself, a ranking approach may be helpful for post-training evaluation. Namely, having trainees produce a correct hierarchy of suitable responses may provide a more thorough assessment of the trainees' comprehension of all the fine points of cultural difference that are applicable to a particular situation. Again, the

merits of a single-preference response vs a hierarchy of response preferences for the evaluation of cultural assimilator training is something that would have to be determined via future research.

Finally, the results of this study potentially also hold implications for the use of performance appraisal forms in multinational corporations. Gregersen, Hite & Black (1996) noted that 76% of U.S. firms used standardized performance appraisal forms across their organisation for the purposes of expatriate evaluation. It seems likely that such standardized assessment approaches are being employed even when PCN managers need to assess an HCN individual's potential (and visa versa) in a geocentric MNC context. Clearly, 7-point Likert scales hold promise for such cross-cultural assessment situations. The ranking approach, however, would seem to be less suitable for international performance appraisal forms, because as noted above, ranking is geared to present respondents with a behavioural situation and then assess the respondent's preference hierarchy among a set of discrete (behavioural) response options. In contrast, performance appraisal forms are typically geared to assess the quality of a target's singular observed behavioural response and to rate that singular response along a continuum.

LIMITATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

Of course our study is not without limitations. First, although we used the same type of questions relating to work values for both studies and only compared the countries included in both studies, the studies were separated in time by 2-5 years and the samples were different: undergraduate students versus MBA students. Hence it is possible that the differences found between the studies are due to effects other than differences in the response format used. On the other hand, the fact that the differences in response style patterns between countries were generally very consistent across the two samples gives us confidence that, for the purposes of our study, the samples are comparable. Moreover, both samples consisted of students studying business, albeit at different levels. However, future studies might want to collect data for both 5-point and 7-point Likert scales within the same study. Of course this would increase the challenges of data collection if respondents are also split with regard to questionnaire language.

Second, although we collected data on (7-point Likert scale) rating and ranking responses within the same study, these data related to slightly different topics. It is possible that the more

meaningful clustering of countries and the weaker language effects for the ranking data was caused by the topic of investigation rather than the response format. However, hierarchical clustering for the ranking data from study 2 showed clear country clusters that generally confirmed earlier research, whilst clusters for the rating data from study 2 were far less differentiated and contradicted clustering results found in previous studies. Moreover, we do not see any intrinsic reason why questions relating to work values should be more susceptible to language effects than questions relating to management styles. In fact, given that the questions on work values could be expected to be related more specifically and more directly to the respondent's daily working life than the more generic scenarios, if anything we would expect response confidence (and hence the absence of language effects) to be higher for the work values questions. However, future researchers might consider using questions dealing with the same content area and differentiating only response format. This would probably necessitate collecting data for the different response options from different respondents as it is unlikely that the same respondents would be willing to both rank and rate the same questions. Even if they would be willing to do so, a desire to appear consistent might result in a similarity that would not be present under normal circumstances.

The final limitation relates to the response format that was shown to perform best in this paper: ranking of solutions. Although we found this response format to perform better in terms of providing meaningful country clusters and a lack of language effect, the statistical tests that can be conducted with ranked data are limited. Popular techniques such as factor analysis and regression analysis can only be conducted with interval data. Hence a very fruitful avenue for further research could be to apply a technique developed in the Marketing literature (Munson & McIntyre, 1979; McCarty & Shrum, 1997, 2000): the most-least rating procedure. This procedure requires respondents to first consider all answer alternatives and rank the most and least important/applicable. After doing so, respondents are then asked to rate each of the remaining alternatives in the same most/least fashion (based on their importance or applicability), and to then continue eliminating pairs of alternatives in this fashion until all possible alternatives are exhausted. This forces the respondents to first consider all of the options and hence increases the likelihood that they rate them in a comparative manner. Research has shown that this technique reduces acquiescence (called end-piling in these studies) and increases differentiation between items (McCarty & Shrum, 1997, 2000), hence enabling the use of

popular statistical analysis techniques such as factor analysis and regression, whilst not increasing the burden on respondents too much. To the best of our knowledge Lenartowicz and Roth (2001) have been the only ones who have used this technique in international management research. We would suggest it merits further research for its applicability in international management as it might well increase data quality considerably.

CLOSING COMMENTS

This article set out to answer the question whether rating or ranking is the best way to reduce response and language bias in cross-national research. Although several articles have investigated differences in response styles between countries and there are some indications that language might influence questionnaire response, so far no large-scale study had been conducted to answer this question. We evaluated two specific solutions to problems caused by cross-country differences in responses styles and the possible effect of language of the questionnaire on the way people respond. The first involved changing the Likert-scale format from the most commonly used 5-point scale to a 7-point scale. The second used ranking rather than rating as a response method. Our results confirmed our hypotheses that both solutions would reduce response and language bias, but that ranking would generally be a superior solution.

We showed that it is possible to ascertain systematic differences between countries that confirm country clusters found in previous research by asking respondents to rank as little as three preferred solutions on 4 scenarios. Scenarios are less sensitive to translation problems as they are less dependent on individual words than short Likert-scale items. Moreover they remove the need for scale anchors that are often very difficult to translate.

Our study showed that even when asking respondents to only rank their top-three preferred alternatives, a task that should not be overly taxing, meaningful results can be achieved. Of course not all research questions might be amenable to being studied by scenarios with ranked solutions. However, we do think that this technique merits a wider application in cross-cultural studies, and we encourage researchers to investigate its use more systematically.

REFERENCES

Allik, J. and 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.

Alwin, D.F. and Krosnick, J.A. (1985) The Measurement of Values in Surveys: A Comparison of Ratings and Rankings, *Public Opinion Quarterly* 49: 535-552.

Baumgartner, Hans and Jan-Benedict E.M. Steenkamp (2001) Response Styles in Marketing Research: A Cross-National Investigation, *Journal of Marketing Research* 38 (May): 143-156.

Bond, M.H. and 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):169-185.

Brislin, R.W. (1989) A culture general assimilator: Preparation for various types of sojourns. *International Journal of Intercultural Relations* 10: 215-234.

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

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

Cushner, K. (1989) Assessing the impact of a culture general assimilator. *International Journal of Intercultural Relations* 13(2):125-146.

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.

Gregersen, H.B., Hite, J.M. and Black, J.S. (1996) Expatriate performance appraisal in U.S. multinational firms. *Journal of International Business Studies*, 27(4): 711-738.

Gupta, V. and Hanges, P.J. (2004) Regional and Climate Clustering of Societal Cultures, in: House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W. and Gupta, V. *Culture, Leadership, and Organizations, The Globe Study of 62 Societies*, Thousand Oaks: Sage, pp.178-218.

Hanges, P.J. and Dickson, M.W. (2004) The Development and Validation of the GLOBE Culture and Leadership Scales, in: House, R.J., Hanges P.J., Javidan, M., Dorfman, P.W. and Gupta, V., *Culture, Leadership, and Organizations, The Globe Study of 62 Societies*, Thousand Oaks: Sage, pp. 122-151.

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 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.

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*. Second Edition, Thousand Oaks: Sage Publications.

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

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

Latham, G.P., Saari, L.M., Pursell, E.D. and Campion, M.A. (1980) The situational interview. *Journal of Applied Psychology*, 65: 422-27.

Lenartowicz, T. and Roth, K. (2001) Does Subculture Within a Country Matter? A Cross-cultural Study of Motivational Domains and Business Performance in Brazil, *Journal of International Business Studies* 32(2): 305-325.

Leung, K. and Bond, M. H. (1989) On the Empirical Investigation of Dimensions for Cross-cultural Comparisons, *Journal of Cross-Cultural Psychology* 20(2): 133-151.

Marin, G., Gamba, R.J. and Marin, B.V. (1992) Extreme Response Style and Acquiescence among Hispanics. The Role of Acculturation and Education', *Journal of Cross-Cultural Psychology* 23(4): 498–509.

McCarty, J.A.; Shrum, L.J. (1997) Measuring the Importance of Positive Constructs: A Test of Alternative Rating Procedures, *Marketing Letters* 8 (2): 239-250.

McCarty, J.A.; Shrum, L.J. (2000) The Measurement of Personal Values in Survey Research. A Test of Alternative Rating Procedures, *Public Opinion Quarterly* 64: 271-298.

Munson, J. M.; McIntyre, S.H. (1979) Developing Practical Procedures for the Measurement of Personal Values in Cross-cultural Marketing, *Journal of Marketing Research* 16: 55-60.

Peng, K.; Nisbett, R.E.; Wong, N.Y.C. (1997) Validity Problems in Comparing Values Across Cultures and Possible Solutions, *Psychological Methods* 2(4): 329-344.

Ralston, D.A., Cunniff, M.K. and Gustafson, D.J. (1995) Cultural Accommodation: The Effect of Language on the Response of Bilingual Hong Kong Chinese Managers, *Journal of Cross-Cultural Psychology* 26 (6): 714-727.

Rokeach, M. (1973) *The nature of human values*, New York: The Free Press.

Ronen, S. and Shenkar, O. (1985) Clustering Countries on Attitudinal Dimensions: A Review and Synthesis, *Academy of Management Review* 10(3): 435-454.

Singh, J. (1995) Measurement Issues in Cross-national Research, *Journal of International Business Studies* 26: 597-620.

Sirota, D. and Greenwood, M. (1971) Understand Your Overseas Workforce, *Harvard Business Review* 49(1): 53-60.

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

Schwartz, S.H. (1999) A Theory of Cultural Values and Some Implications for Work, *Applied Psychology: An International Review*, 48(1): 23–47.

Takahashi, K., Ohara, N., Antonucci, T.C. and 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.

Tsui, A.S.; Nifadkar, S.S.; Ou, A.Y. (2007) Cross-National, Cross-Cultural Organizational Behavior Research: Advances, Gaps, and Recommendations *Journal of Management* 2007; 33: 426-478

Usunier (1998). *International and cross-cultural management research*. London: Sage.

Van de Vijver, F.J.R. and Leung, K. (2000) Methodological Issues in Psychological Research on Culture, *Journal of Cross-Cultural Psychology* 31: 33-51.

Voss, K.E.; Stem, D.E.; Johnson, L.W. and Arce, C. (1996) An Exploration of the Comparability of Semantic Adjectives in Three Languages. A Magnitude Estimation Approach, *International Marketing Review* 13(5): 44-58.

Appendix 1: Example of a scenario used in the study

Imagine you are a manager in a Chilean (varied by country) company that produces a high-technology product. You and one of your superiors are attending a meeting with potential clients. You have a very good knowledge of the technical aspects of the product that your company sells, because of your previous job experience as a technical engineer. During the meeting, your superior makes a mistake in describing the features of the product, because he doesn't know too much about technical issues. There is no way to inform your superior of his mistake during the meeting without the clients noticing it. What would **you** do? Please rank **the best three alternatives** from 1 to 3.

	Politely correct your superior in the meeting.
	Pretend to be responsible for the mistake yourself.
	Mention the correct features in the meeting without referring to your superiors earlier description.
	Say nothing in the meeting, but talk to your superior afterwards, so that he can decide on a way to inform the clients of his mistake.
	Say nothing in the meeting, but arrange for the clients to receive full technical information afterwards. In that way the clients can verify the details themselves.
	Do nothing. It is not your responsibility to give the clients technical information.
	Do nothing. Any action you take would make your superior lose face.

Table 2: Proximity matrix for management scenarios, ranking,

Proximity Matrix

Rescaled Squared Euclidean Distance

Case	1:Brazil	2:Canada	3:Chile	4:Finland	5:Germany	6:Greece	7:India	8:Ireland	9:Lithuani	10:Malaysia	11:Mexico	12:Netherla	13:Philippi	14:Portugal	15:Sweden	16:Thailand	17:Taiwan	18:Turkey	19:UK	20:USA
1:Brazil	.000	.125	.213	.324	.174	.269	.657	.156	.610	.423	.105	.263	.404	.004	.238	.646	.547	.526	.228	.169
2:Canada	.125	.000	.276	.157	.055	.440	.608	.024	.667	.431	.228	.182	.398	.116	.160	.481	.549	.611	.036	.045
3:Chile	.213	.276	.000	.330	.266	.547	.691	.288	.543	.320	.156	.316	.508	.170	.491	.468	.575	.632	.241	.216
4:Finland	.324	.157	.330	.000	.181	.776	.851	.224	.943	.725	.332	.539	.853	.400	.466	.684	.953	.944	.129	.230
5:Germany	.174	.055	.266	.181	.000	.557	.797	.081	.791	.623	.289	.129	.607	.161	.189	.719	.594	.726	.139	.098
6:Greece	.269	.440	.547	.776	.557	.000	.618	.356	.404	.467	.443	.575	.532	.291	.703	.498	.519	.408	.703	.374
7:India	.657	.608	.691	.851	.797	.618	.000	.662	.382	.299	.891	.936	.537	.786	1.000	.497	.912	.569	.792	.863
8:Ireland	.156	.024	.288	.224	.081	.356	.662	.000	.576	.529	.185	.220	.575	.185	.213	.457	.321	.392	.075	.000
9:Lithuani	.610	.667	.543	.943	.791	.404	.382	.576	.000	.181	.606	.783	.619	.623	.869	.555	.750	.237	.689	.597
10:Malaysia	.423	.431	.320	.725	.623	.467	.299	.529	.181	.000	.495	.570	.307	.395	.745	.555	.894	.470	.556	.522
11:Mexico	.105	.228	.156	.332	.289	.443	.891	.185	.606	.495	.000	.311	.559	.149	.296	.472	.446	.506	.143	.202
12:Netherla	.263	.182	.316	.539	.129	.575	.936	.220	.783	.570	.311	.000	.504	.220	.212	.754	.512	.685	.301	.225
13:Philippi	.404	.398	.508	.853	.607	.532	.537	.575	.619	.307	.559	.504	.000	.328	.797	.652	.783	.892	.644	.594
14:Portugal	.004	.116	.170	.400	.161	.291	.786	.185	.623	.395	.149	.220	.328	.000	.253	.666	.510	.663	.238	.098
15:Sweden	.238	.160	.491	.466	.189	.703	1.000	.213	.869	.745	.296	.212	.797	.253	.000	.780	.476	.564	.197	.290
16:Thailand	.646	.481	.468	.684	.719	.498	.497	.457	.555	.555	.472	.754	.652	.666	.780	.000	.408	.472	.486	.538
17:Taiwan	.547	.549	.575	.953	.594	.519	.912	.321	.750	.894	.446	.512	.783	.510	.476	.408	.000	.300	.591	.486
18:Turkey	.526	.611	.632	.944	.726	.408	.569	.392	.237	.470	.506	.685	.892	.663	.564	.472	.300	.000	.647	.575
19:UK	.228	.036	.241	.129	.139	.703	.792	.075	.689	.556	.143	.301	.644	.238	.197	.486	.591	.647	.000	.101
20:USA	.169	.045	.216	.230	.098	.374	.863	.000	.597	.522	.202	.225	.594	.098	.290	.538	.486	.575	.101	.000

This is a dissimilarity matrix

Table 3: Proximity matrix for work values, rating,

Proximity Matrix

Case		Rescaled Squared Euclidean Distance																		
	1:Brazil	2:Canada	3:Chile	4:Finland	5:Germany	6:Greece	7:India	8:Ireland	9:Lithuani	10:Malaysia	11:Mexico	12:Netherla	13:Philippi	14:Portugal	15:Sweden	16:Thailand	17:Taiwan	18:Turkey	19:UK	20:USA
1:Brazil	.000	.042	.198	.327	.444	.184	.354	.129	.246	.170	.141	.269	.227	.009	.260	.218	.174	.272	.259	.092
2:Canada	.042	.000	.208	.150	.203	.210	.371	.028	.314	.251	.234	.108	.344	.016	.073	.238	.189	.365	.095	.095
3:Chile	.198	.208	.000	.558	.746	.230	.487	.330	.425	.170	.060	.539	.207	.222	.497	.140	.253	.203	.597	.168
4:Finland	.327	.150	.558	.000	.018	.471	.659	.162	.546	.624	.588	.027	.818	.187	.055	.616	.432	.821	.056	.485
5:Germany	.444	.203	.746	.018	.000	.609	.776	.253	.641	.779	.750	.044	.957	.272	.093	.762	.564	1.000	.069	.605
6:Greece	.184	.210	.230	.471	.609	.000	.186	.192	.250	.052	.144	.483	.079	.222	.434	.110	.132	.175	.398	.063
7:India	.354	.371	.487	.659	.776	.186	.000	.408	.318	.199	.421	.603	.223	.383	.582	.225	.203	.264	.563	.268
8:Ireland	.129	.028	.330	.162	.253	.192	.408	.000	.353	.294	.352	.117	.403	.135	.130	.308	.203	.435	.049	.099
9:Lithuani	.246	.314	.425	.546	.641	.250	.318	.353	.000	.202	.442	.577	.317	.292	.592	.330	.223	.246	.507	.252
10:Malaysia	.170	.251	.170	.624	.779	.052	.199	.294	.202	.000	.107	.602	.000	.263	.563	.018	.075	.036	.572	.039
11:Mexico	.141	.234	.060	.588	.750	.144	.421	.352	.442	.107	.000	.549	.099	.175	.500	.124	.222	.165	.595	.144
12:Netherla	.269	.108	.539	.027	.044	.483	.603	.117	.577	.602	.549	.000	.744	.149	.029	.585	.390	.781	.024	.434
13:Philippi	.227	.344	.207	.818	.957	.079	.223	.403	.317	.000	.099	.744	.000	.329	.716	.028	.135	.021	.712	.089
14:Portugal	.009	.016	.222	.187	.272	.222	.383	.135	.292	.263	.175	.149	.329	.000	.134	.292	.207	.366	.172	.183
15:Sweden	.260	.073	.497	.055	.093	.434	.582	.130	.592	.563	.500	.029	.716	.134	.000	.562	.388	.721	.060	.379
16:Thailand	.218	.238	.140	.616	.762	.110	.225	.308	.330	.018	.124	.585	.028	.292	.562	.000	.080	.033	.576	.084
17:Taiwan	.174	.189	.253	.432	.564	.132	.203	.203	.223	.075	.222	.390	.135	.207	.388	.080	.000	.115	.379	.124
18:Turkey	.272	.365	.203	.821	1.000	.175	.264	.435	.246	.036	.165	.781	.021	.366	.721	.033	.115	.000	.745	.114
19:UK	.259	.095	.597	.056	.069	.398	.563	.049	.507	.572	.595	.024	.712	.172	.060	.576	.379	.745	.000	.342
20:USA	.092	.095	.168	.485	.605	.063	.268	.099	.252	.039	.144	.434	.089	.183	.379	.084	.124	.114	.342	.000

This is a dissimilarity matrix

Figure 1: Dendrogram for management scenarios, ranking,

▽

* * * * * * H I E R A R C H I C A L   C L U S T E R   A N A L Y S I S * * * * * *

Dendrogram using Ward Method

```
                     Rescaled Distance Cluster Combine

    C A S E            0         5        10        15        20        25
  Label        Num     +---------+---------+---------+---------+---------+

  USA           19     ─┬─┐
  Ireland       20     ─┘ ├─┐
  Canada         2     ─┬─┘ ├─┐
  UK            18     ─┘   │ ├─┐
  Germany        5     ─────┘ │ │
  Finland        4     ───────┘ ├─────────────────────────────────────┐
  Netherlands   11     ─┬───────┘                                      │
  Sweden        14     ─┘                                              │
  Brazil         1     ─┬─┐                                            │
  Portugal      13     ─┘ ├─┐                                          │
  Chile          3     ───┘ │                                          │
  Mexico        10     ─────┘                                          │
  Taiwan        16     ─┬───┐                                          │
  Turkey        17     ─┘   ├───┐                                      │
  Thailand      15     ─────┘   ├───────────┐                          │
  Greece         6     ─────────┘           ├──────────────────────────┘
  Lithuania      8     ─┬─┐                 │
  Malaysia       9     ─┘ ├─┐               │
  India          7     ───┘ ├───────────────┘
  Philippines   12     ─────┘
```
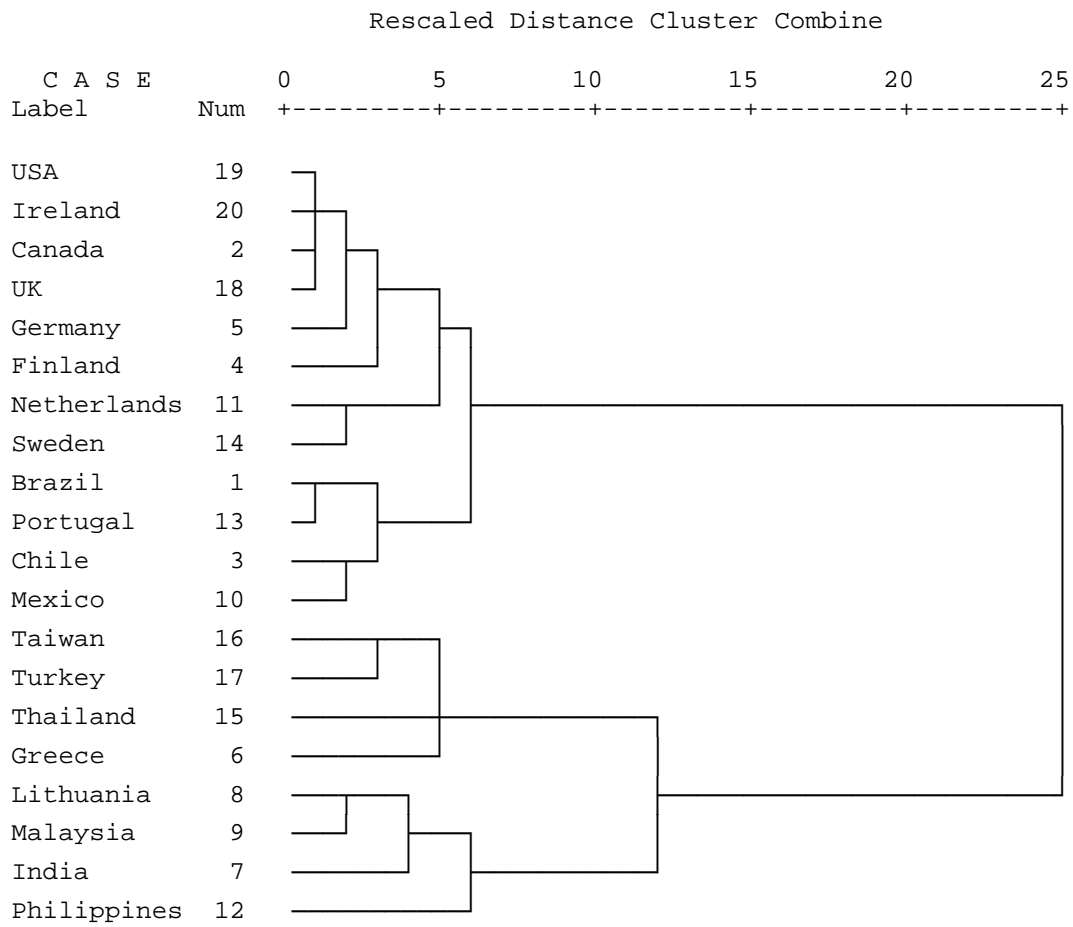
30

Figure 2: Dendrogram for work values, rating,

▽

* * * * * * H I E R A R C H I C A L   C L U S T E R   A N A L Y S I S * * * * * *

Dendrogram using Ward Method

                        Rescaled Distance Cluster Combine

    C A S E   0   5   10   15   20   25
  Label   Num   +---------+---------+---------+---------+---------+

  Malaysia   9   ─┐
  Philippines   12   ─┤
  Thailand   15   ─┼─┐
  Turkey   17   ─┘ │
  Taiwan   16   ─┐ ├─┐
  Greece   6   ─┼─┘ │
  USA   19   ─┘   ├─┐
  Chile   3   ─┐   │ │
  Mexico   10   ─┼─┐ │ │
  India   7   ─┘ ├─┘ ├──────────────────────────────────────┐
  Lithuania   8   ───┘   │   │
  Finland   4   ─┐   │   │
  Germany   5   ─┼─┐   │   │
  Netherlands   11   ─┤ ├───┘   │
  UK   18   ─┤ │   │
  Sweden   14   ─┘ │   │
  Brazil   1   ─┐ │   │
  Portugal   13   ─┼─┤   │
  Canada   2   ─┤ │   │
  Ireland   20   ─┘