
RANKING AND RATING IN NATIVE-LANGUAGE VERSUS ENGLISH-LANGUAGE QUESTIONNAIRES: A METHODOLOGICAL COMPARISON

Anne-Wil HARZING, Department of Management, University of Melbourne, Australia*

Joyce BALDUEZA, Worldbank, Philippines¹

Wilhelm BARNER-RASMUSSEN, Swedish School of Economics, Finland

Cordula BARZANTNY, Groupe ESC Toulouse Business School, France

Anne BERTHELOT, University of Texas at El Paso, 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

Sharon LEIBA O'SULLIVAN, Telfer School of Management, University of Ottawa, Canada

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

Paulo PROCHNO, Fundação Dom Cabral, Brazil

Srabani ROY CHOUDHURY, International Management Institute, India

Ayşe SAKA, University of Groningen, Netherlands

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

* Corresponding author. Anne-Wil HARZING, Department of Management, University of Melbourne, Parkville Campus, Parkville, Victoria 3010, Australia, Tel: (61) (3) 8344 3724, Fax: (61) (3) 9349 4293, harzing@unimelb.edu.au

¹ Co-authors 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.

RANKING AND RATING IN NATIVE-LANGUAGE VERSUS ENGLISH-LANGUAGE QUESTIONNAIRES: A METHODOLOGICAL COMPARISON

ABSTRACT

Cross-national research is plagued by many problems not present in purely domestic research. Among these are country differences in responses styles and the potential of the language of the questionnaire to affect the way people respond. In this article we present a methodological comparison of two different response style formats – ranking and rating – and assess the extent to which they can help to address the problems of response style differences and language effects. For rating we assess the effect of changing the commonly used 5-point scales to 7-point scales. For ranking we present respondents with short scenarios for which they need to rank their top 3 solutions. Our results - based on two studies of undergraduate and MBA students in 16 different countries - confirm our hypotheses that both solutions reduce response and language bias, but that ranking generally is a superior solution.

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 and Leung, 2000).

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 and 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 rather than picking up real differences in the management phenomena across countries. Research has also shown that the use of English-language questionnaires might obscure important differences between countries through a reduced variance in responses, caused either by cultural accommodation or by a lack of the respondent's confidence in responding in a non-native language.

This article evaluates two solutions to the problems caused by cross-country differences in response styles and the effect of the language on questionnaire response. 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. We hypothesise that both solutions can reduce response and language bias, but that ranking will generally provide a superior solution in this respect. Our results, based on two studies conducted with a sample of 1801 undergraduate and 1675 MBA students in 16 countries generally confirm our hypotheses.

The remainder of this article is structured as follows. In order to substantiate our hypotheses, the second section reviews the literature on response styles and language effects in cross-cultural research. Subsequently, the third section describes our sample, measures and method of analysis. After the fourth section has presented the findings of our empirical study, the fifth closes with a discussion and conclusion.

LITERATURE REVIEW

RESPONSE STYLES

A large number of studies have confirmed that there are substantial and systematic differences in response styles between countries (for 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. Country-level extraversion was also shown to impact on both positive ERS and ARS.

One solution that has been proposed 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. Hui and Triandis (1989) found that ERS for Hispanics disappeared when 10-point Likert scales were used. In addition, we argue that increasing the number of answer alternatives on a Likert scale might 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 seven-point Likert scales will include a lower proportion of middle response styles and extreme response styles than responses to five-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) with an option to express a relatively strong (dis)agreement without having to resort to the scale extremes. It also pro-

vides respondents from countries with a higher incidence of MRS (e.g. East Asian countries) 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 middle response styles and extreme response styles will be less pronounced for 7-point Likert scales than for 5-point Likert scales.

Whereas increasing the number of answer alternatives might alleviate the incidence of MRS and ERS, it does nothing to reduce ARS or DRS. Moreover, appropriate translation of scale anchors into other languages is often difficult. Even if they do translate into appropriate local equivalents, the intensity associated with these equivalents may be different from the original language.² Instead of having scale anchors reflect levels of importance or (dis)agreement, they can be incorporated into the question and reflect opposites. Some of the items that were used in the Globe study (House et al. 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 carefully as most scale anchors would be different, 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 as words that are seen as opposites in some countries might not be opposites in other countries.

In this study, we therefore used another remedy, namely to ask respondents to rank statements rather than using Likert scales. This technique has been used in some studies comparing cul-

tural 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 scenarios with ranked solutions will result in more well-defined differences across countries than comparisons based on ratings of statements using Likert scales. Hence:

Hypothesis 2: Country clusters based on ranked responses will be more well-defined than country clusters based on Likert scale responses.

IMPACT OF LANGUAGE

A second problem that has been identified in doing cross-national research is the possibility that the language of the questionnaire might impact on 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). Some studies have identified cultural accommodation as a reason for responses 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.

² A particularly striking example is provided by Voss, Stem, Johnson & Arce (1996) who show that while the magnitude

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

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 MRS (ERS) (Harzing, 2006) supports this assumption. Since in comparison to 7-point Likert scales, 5-point Likert scales provide a smaller range of relatively neutral and relatively extreme responses, the reduction in variance for English-language questionnaires in comparison to native-language questionnaires could be expected to be greater for 5-point Likert scales.

Hypothesis 3b: The reduction of between-country differences when comparing native-language questionnaires with English-language questionnaires will be greater for 5-point Likert scales than for 7-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 enhanced by the context provided in the scenario and the fact that the solution statements are generally longer than statements used in Likert scales. Hence, we expect the reduction in variance between the native-language version and the English-language version to be greater for Likert scale questions than for rankings.

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

estimates for good and very good were 74 and 87 in English, they were 91 and 101 in the equivalent Japanese translation.

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. 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. Some countries were dropped and others added in the second project. 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.³ In the tests that only compare 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

³ 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.

from our sample, so that our comparisons only included students who could be assumed to be representative of the country they studied in. Although data were collected on a voluntary basis response rates were high, generally between 80-100%. The resulting sample sizes ranged from 46 for the UK, where only English-language questionnaires were used, to 147 for Portugal, but for most countries were around 100. Data were collected in-class 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. 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 37% female respondents. Again, international students were excluded from our sample and response rates were in the order of 80-100%. The resulting sample sizes ranged from 41 for the Philippines to 168 for Portugal, but for most countries were around 100. Data were collected 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 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). It is also important to note that the universities included in our study were generally public universities and hence their students might be expected to be more representative of the general population than students in private universities. For those countries where data were collected at private universities, local collaborators indicated that they had no reason to believe that their students represented a different sub-section of society than in public universities.

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 and less orthogonal than 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 or 4) 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.

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 sen-

tences were not felt to be fully equivalent a better translation was sought through discussion between the three or four 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 and Chinese) took several sessions lasting up to eight hours in total.

Although, as in any multi-country study, it is very difficult to guarantee translation accuracy with absolute certainty, we are quite confident that the resulting questionnaires are equivalent in meaning across languages. Further, any potentially remaining translation inaccuracies would be attenuated by two factors. First, we will be looking at response patterns for a total of 18-29 items and hence translation inaccuracies in one item would not have a major impact on overall results. Scenarios are even less sensitive to translation inaccuracies 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. Second, we are looking at overall response patterns across 16 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 counties in the 1st study or the 18 countries in the 2nd study showed a significant difference between 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 and only differed from each other in terms of the language.

ANALYSIS

Oneway 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

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. As expected all countries showed a reduction in MRS and overall the proportion of middle responses changed from 21.9% to 13.6% (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). Most countries also showed a reduction in ERS; only Malaysia, the USA and Turkey showed an increase in ERS. This increase was marginal for Malaysia and the USA, but substantial (from 18.1% to 28.4%) for Turkey. Overall ERS declined from 27.4% to 22.3%. Hence, we find confirmation for hypothesis 1a. The picture for ARS and DRS was more mixed. As expected the overall ARS and DRS increased, however some individual countries deviated from this pattern. The balance be-

tween ARS and DRS, which by some authors is seen as the best measure of acquiescence bias decreased slightly, with Turkey and Malaysia being notable exceptions in showing a strong increase in acquiescence bias.

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	21.9%	27.4%	67.7%	9.8%	57.9%	9.625 ***	15.505 ***	21.944 ***	27.236 ***
Work values 2006 7-point scale	13.6%	22.3%	71.1%	14.7%	56.4%	4.283 ***	10.867 ***	19.047 ***	24.216 ***

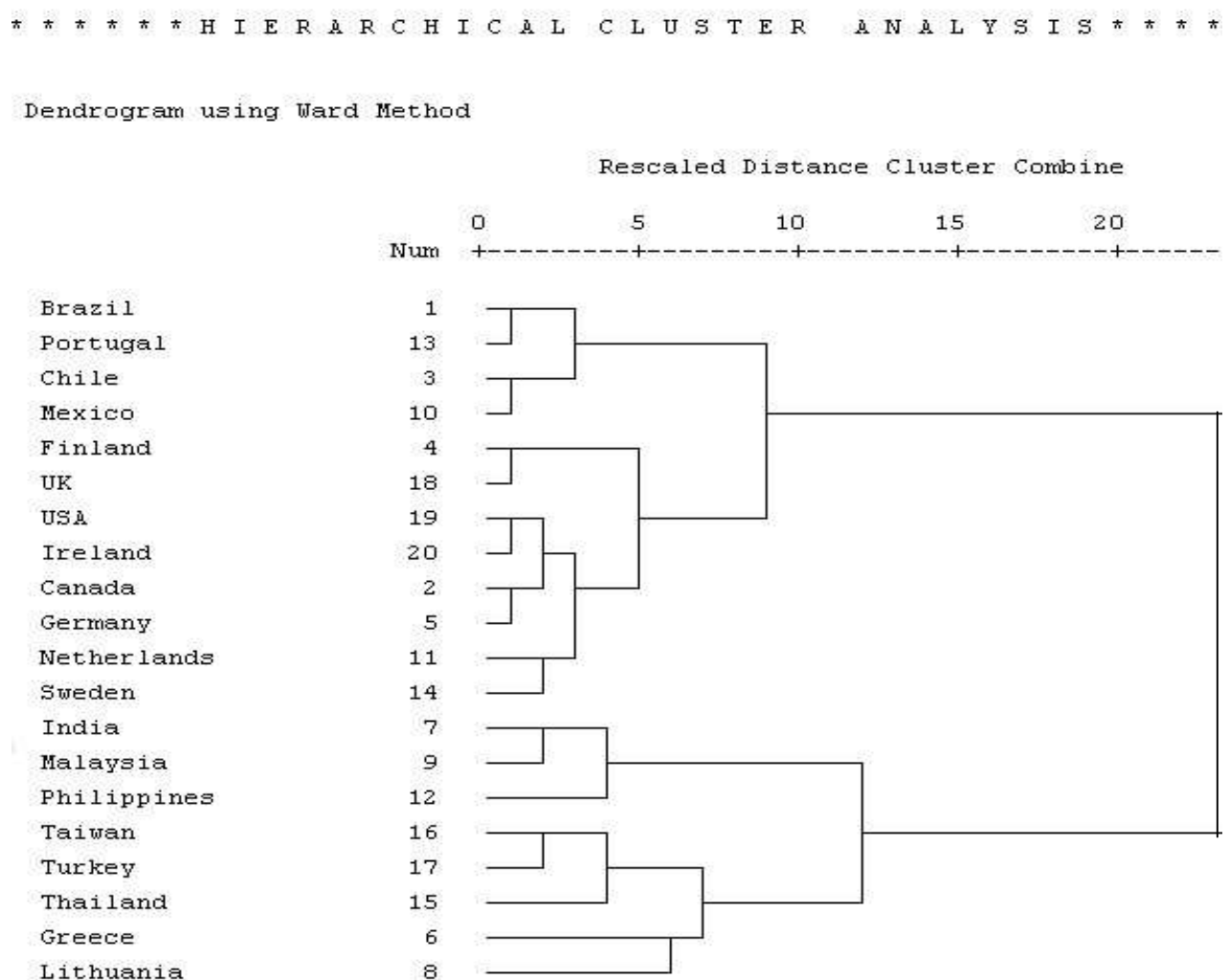
When comparing response styles between countries (hypothesis 1b), the variance for MRS and ERS was reduced and hence countries are more similar in this respect when a 7-point scale is used. This confirms hypothesis 1b. Country differences in ARS and DRS remained very similar. Overall the country patterns in responses styles are *very* similar between the two studies. The Latin American countries (in particular Chile & Mexico), India and Malaysia show low MRS, low DRS, and high ARS and ERS, whereas the Western European countries show high MRS, high DRS and low ARS and ERS.

CLUSTERING COUNTRIES

To test hypothesis 2 and illustrate the advantages of ranking over rating, we conducted a hierarchical cluster analysis for both the scenario data and the work values data in the second study. 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 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) – form a slightly more heterogenous mix of East Asian, South-East Asian and Cen-

tral/Eastern European countries. The latter group though includes many countries that would be seen as emerging economies. 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. Most of these clusters have been well-established in previous cross-cultural studies such as those by Hofstede (1980, 2001) and Ronen & Shenkar (1985).

Figure 1: Dendrogram for management scenarios (ranking)



The accompanying proximity matrix (Table 2) quantifies the distance between the individual countries on a scale from 0 to 1 and again confirms established knowledge in this area. Brazil's clos-

est 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 form a tight cultural cluster. Overall, Lithuania shows the largest distance from the other countries in the sample and is particularly distant from the Northern European countries, its closest neighbour being Turkey. This confirms an earlier study (Harzing, 2004) that showed it to cluster with Turkey and the other Eastern European countries (Bulgaria, Poland, Russia).

Table 2: Proximity matrix for management scenarios (ranking)

Proximity Matrix																				
	Rescaled Squared Euclidean Distance																			
	Brazil	Canada	Chile	Finland	Germany	Greece	India	Ireland	Lithuania	Malaysia	Mexico	Netherlands	Philippines	Portugal	Sweden	Thailand	Taiwan	Turkey	UK	USA
Brazil	.00	.16	.16	.44	.21	.19	.45	.16	.53	.30	.07	.31	.22	.00	.23	.42	.40	.38	.32	.14
Canada	.16	.00	.28	.17	.03	.51	.59	.04	.80	.38	.21	.14	.24	.16	.13	.45	.61	.71	.05	.05
Chile	.16	.28	.00	.31	.35	.58	.38	.26	.47	.24	.10	.38	.34	.20	.39	.28	.41	.48	.28	.18
Finland	.44	.17	.31	.00	.34	.94	.96	.27	.99	.74	.24	.40	.73	.46	.25	.65	.99	.96	.08	.22
Germany	.21	.03	.35	.34	.00	.54	.79	.10	.88	.50	.36	.09	.37	.18	.16	.70	.63	.76	.24	.11
Greece	.19	.51	.58	.94	.54	.00	.51	.38	.51	.43	.45	.64	.52	.31	.67	.55	.57	.38	.88	.41
India	.45	.59	.38	.96	.79	.51	.00	.62	.40	.12	.58	.89	.40	.56	.93	.36	.78	.51	.78	.66
Ireland	.16	.04	.26	.27	.10	.38	.62	.00	.62	.45	.19	.22	.37	.20	.19	.36	.32	.42	.13	.01
Lithuania	.53	.80	.47	.99	.88	.51	.40	.62	.00	.30	.49	1.00	.70	.63	.86	.49	.64	.24	.79	.52
Malaysia	.30	.38	.24	.74	.50	.43	.12	.45	.30	.00	.37	.52	.17	.29	.62	.37	.70	.43	.58	.41
Mexico	.07	.21	.10	.24	.36	.45	.58	.19	.49	.37	.00	.39	.34	.13	.22	.41	.46	.47	.17	.10
Netherlands	.31	.14	.38	.40	.09	.64	.89	.22	1.00	.52	.39	.00	.38	.31	.18	.66	.70	.81	.27	.15
Philippines	.22	.24	.34	.73	.37	.52	.40	.37	.70	.17	.34	.38	.00	.22	.60	.53	.73	.78	.44	.36
Portugal	.00	.16	.20	.46	.18	.31	.56	.20	.63	.29	.13	.31	.22	.00	.27	.49	.46	.57	.36	.12
Sweden	.23	.13	.39	.25	.16	.67	.93	.19	.86	.62	.22	.18	.60	.27	.00	.65	.53	.51	.17	.18
Thailand	.42	.45	.28	.65	.70	.55	.36	.36	.49	.37	.41	.66	.53	.49	.65	.00	.30	.30	.46	.38
Taiwan	.40	.61	.41	.99	.63	.57	.78	.32	.64	.70	.46	.70	.73	.46	.53	.30	.00	.19	.69	.42
Turkey	.38	.71	.48	.96	.76	.38	.51	.42	.24	.43	.47	.81	.78	.57	.51	.30	.19	.00	.75	.54
UK	.32	.05	.28	.08	.24	.88	.78	.13	.79	.58	.17	.27	.44	.36	.17	.46	.69	.75	.00	.11
USA	.14	.05	.18	.22	.11	.41	.66	.01	.52	.41	.10	.15	.36	.12	.18	.38	.42	.54	.11	.00

This is a dissimilarity matrix

In contrast to earlier studies, there is a rather small difference between the USA and the Latin American countries, in particular Mexico. 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 development of the Latin American countries and increased economic interaction through agreements such as NAFTA.

The dendrogram produced based on the work values data (Figure 2) shows far less differentiation. Moreover, countries that in other studies (including the scenario data in the same study) have been clearly identified as being similar – such as the four Anglo and the three Latin American countries & Portugal – are separated into very different clusters. This might 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, Thailand and Mexico. Harzing (2006) already showed that the USA had an acquiescence balance (ARS – DRS) which was much lower than that of Northern and Western European countries and similar in magnitude to that of the Southern European countries. The fact that in our study, the USA even shows stronger acquiescence tendencies that approximate the Asian and Latin European countries might be caused by the location of data collection on the Mexican border. Lithuania is now far less different from the other countries than it is in the scenario study, although its most distant neighbours are still the Northern European countries. In fact it is now The Netherlands and Germany, and to a lesser extent Sweden, Finland and the UK, that are most different from other countries because of their distinct response style, i.e. a relatively low acquiescence and high disacquiescence. As a result, we find confirmation of hypothesis 2: country clusters are more well-defined when using ranking than when using rating.

Figure 2: Dendrogram for work values (rating)

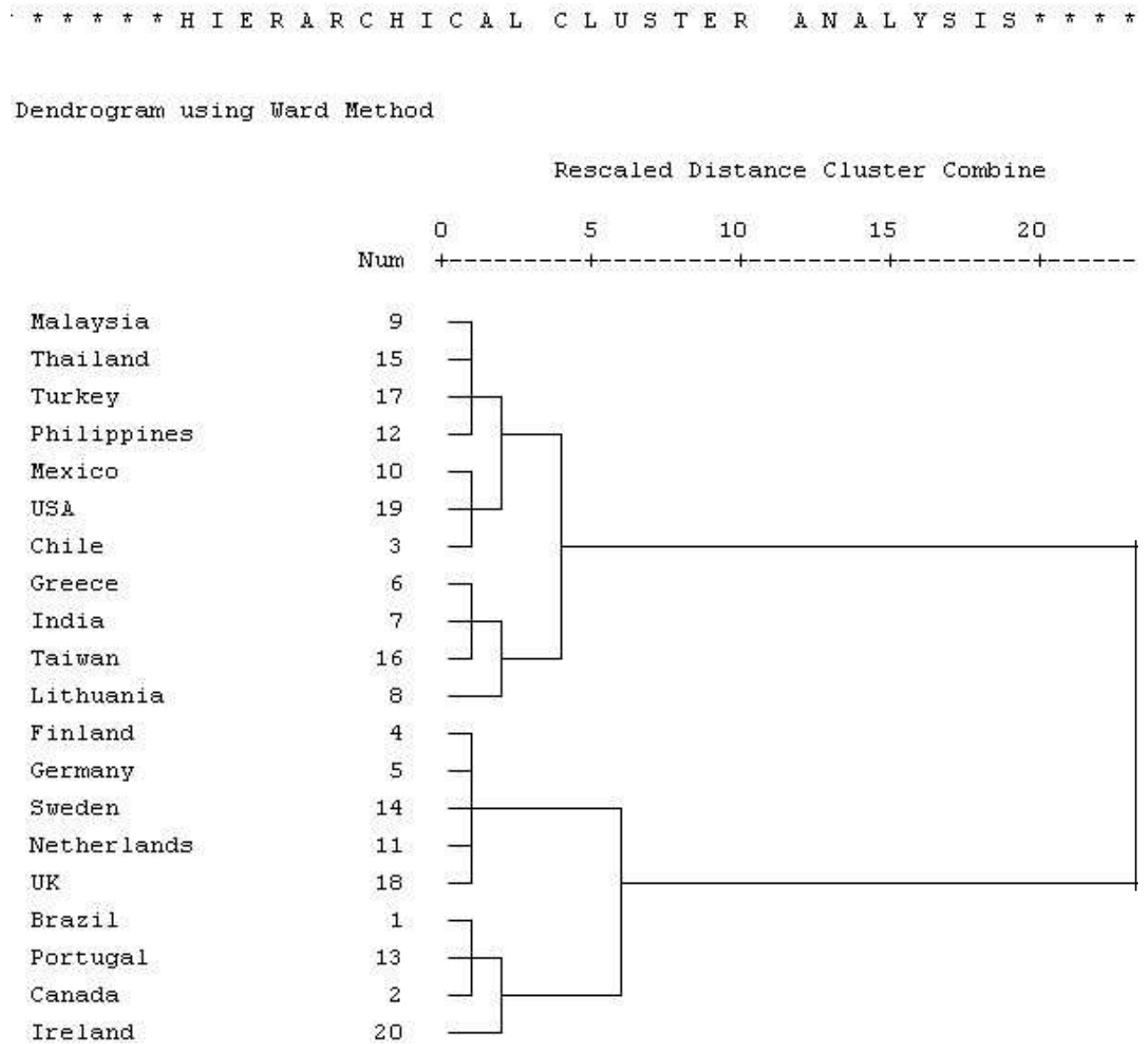


Table 3: Proximity matrix for work values (rating)

Proximity Matrix																				
	Rescaled Squared Euclidean Distance																			
	Brazil	Canada	Chile	Finland	Germany	Greece	India	Ireland	Lithuania	Malaysia	Mexico	Netherlands	Philippines	Portugal	Sweden	Thailand	Taiwan	Turkey	UK	USA
Brazil	.00	.04	.15	.32	.35	.13	.15	.12	.19	.15	.11	.37	.26	.02	.31	.16	.16	.18	.29	.09
Canada	.04	.00	.25	.16	.16	.17	.17	.04	.27	.24	.21	.19	.41	.02	.12	.25	.20	.31	.13	.16
Chile	.15	.25	.00	.69	.71	.25	.22	.39	.29	.14	.07	.76	.16	.18	.66	.16	.30	.17	.71	.11
Finland	.32	.16	.69	.00	.01	.43	.43	.18	.57	.63	.56	.03	.91	.19	.02	.62	.41	.73	.06	.60
Germany	.35	.16	.71	.01	.00	.46	.45	.19	.55	.65	.61	.06	.93	.23	.05	.63	.45	.78	.06	.61
Greece	.13	.17	.25	.43	.46	.00	.10	.20	.13	.06	.13	.53	.15	.16	.45	.09	.12	.14	.40	.07
India	.15	.17	.22	.43	.45	.10	.00	.26	.16	.08	.18	.49	.16	.14	.45	.06	.10	.10	.43	.13
Ireland	.12	.04	.39	.18	.19	.20	.26	.00	.29	.31	.31	.17	.51	.15	.15	.34	.23	.39	.09	.21
Lithuania	.19	.27	.29	.57	.55	.13	.16	.29	.00	.13	.31	.67	.27	.26	.61	.16	.13	.15	.52	.15
Malaysia	.15	.24	.14	.63	.65	.06	.08	.31	.13	.00	.08	.72	.03	.21	.64	.00	.09	.01	.61	.02
Mexico	.11	.21	.07	.56	.61	.13	.18	.31	.31	.08	.00	.65	.09	.13	.57	.11	.22	.12	.57	.06
Netherlands	.37	.19	.76	.03	.06	.53	.49	.17	.67	.72	.65	.00	1.00	.25	.05	.69	.47	.82	.05	.68
Philippines	.26	.41	.16	.91	.93	.15	.16	.51	.27	.03	.09	1.00	.00	.34	.92	.04	.21	.04	.88	.07
Portugal	.02	.02	.18	.19	.23	.16	.14	.15	.26	.21	.13	.25	.34	.00	.19	.21	.19	.26	.24	.18
Sweden	.31	.12	.66	.02	.05	.45	.45	.15	.61	.64	.57	.05	.92	.19	.00	.65	.47	.74	.05	.56
Thailand	.16	.25	.16	.62	.63	.09	.06	.34	.16	.00	.11	.69	.04	.21	.65	.00	.10	.01	.61	.06
Taiwan	.16	.20	.30	.41	.45	.12	.10	.23	.13	.09	.22	.47	.21	.19	.47	.10	.00	.11	.41	.15
Turkey	.18	.31	.17	.73	.78	.14	.10	.39	.15	.01	.12	.82	.04	.26	.74	.01	.11	.00	.70	.05
UK	.29	.13	.71	.06	.06	.40	.43	.09	.52	.61	.57	.05	.88	.24	.05	.61	.41	.70	.00	.52
USA	.09	.16	.11	.60	.61	.07	.13	.21	.15	.02	.06	.68	.07	.18	.56	.06	.15	.05	.52	.00

This is a dissimilarity matrix

IMPACT OF LANGUAGE

We hypothesised that differences *between* countries would be larger for the native language version of the questionnaire items than for the English-language versions. In order to test this we ran a split-sample 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. As Table 4 shows our results confirmed hypothesis 3a. In each of the three instances the overall average F-value or Chi-square statistic was higher in the native version compared to the English-language version.

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	13.31	7.24	46%	72%
Work values 2006	10.14	6.94	32%	69%
Scenarios 2006	60.63	42.11	31%	25%

In hypothesis 3b we argued that using 5-point Likert scales would lead to a stronger reduction in variance for the English-language version than using 7-point Likert scales. 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”, “Be able to serve your country”. Hence we find only partial confirmation of hypothesis 3b.

The reason for expecting the reduction in variance for the English-language version of the 7-point Likert scale to be weaker than for the 5-point Likert scale, was that 7-point scales would be less likely to induce respondents to choose the middle response for the English-language version or an extreme response for the native language version. If this reason is valid, the difference between the language versions in terms of MRS and ERS should be smaller for the 7-point Likert scale than for the 5-point Likert scale. Table 5 shows that this is indeed the case; whereas the difference is in the same direction for both studies, it is only significant for the study using the 5-point Likert scale.

Finally, in hypothesis 3c we expected ratings to show a stronger reduction in variance for the English-language version than rankings. As Table 4 shows this is clearly 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, the proportion of items that has a reduction significant at 0.01 is substantially lower for rankings. Hence, we find confirmation for hypothesis 3c.

Table 5: Differences in response styles between native-language and English-language versions

Type of ques- tions	MRS Native	MRS English	Difference	ERS Native	ERS English	Difference
Work values 2002 5-point scale	20.7%	23.2%	t=3.920***	29.0%	24.3%	t=4.593***
Work values 2006 7-point scale	13.1%	13.7%	t=1.282	23.5%	22.2%	t=1.389

We argued that any reduced variance between countries for the scenarios might be caused by providing less consistent responses in the foreign language. As Table 4 showed this reduced variance was generally very modest. When we compared the standard deviation of responses for the native language version and the English language version, we indeed found the latter to be slightly higher (0.890 versus 0.875). However, this difference was not significant ($t=-0.645$, $p=0.524$). Most countries showed similar standard deviations in both language versions. Only Brazil and Chile and to a much lesser extent Mexico, Portugal, Turkey and Taiwan showed higher standard deviations in the English-language version.

DISCUSSION

Our results showed that even though changing from a 5-point Likert scale to a 7-point Likert scale reduced MRS and ERS, 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 (Leung and Bond, 1989). 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. 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 formed based on the ranked scenario solutions provided meaningful results, clustering based on the rated work values questions mainly reflected differences in responses 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. In each of the three instances the overall average F-value or Chi-square statistic was higher in the native version compared to the English-language version. However, the reduction in variance was smaller for 7-point Likert scales than it was for 5-point Likert scales. We expected that the reason for this was that the 7-point scale would be less likely to induce respondents to choose the middle response for the English-language version or an extreme response for the native language version and found that this was indeed the case. Whereas for the 5-point Likert scales there was a significant difference between language versions in this respect, this was not true for the 7-point Likert scale. 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. This was further supported by the fact that standard deviations were only slightly higher in English than in the native-language version.

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 should allow researchers to have greater confidence in the validity of cross-national differences if these response alternatives are used rather than 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.

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 four 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. 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 data collection challenge 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 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 showed clear country clusters that generally confirmed earlier research, whilst clusters for the rating data 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 re-

sponse 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 alternatives in the usual fashion. This forces 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), whilst not increasing the burden on respondents too much. To the best of our knowledge Lenartowicz & Roth (2001) have been the only ones who have used this technique in international management research. We would suggest it merits further research as it might well increase data quality considerably.

CONCLUSION

In this article we evaluated two solutions to problems caused by cross-country differences in responses styles and the 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 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. & 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*, Vol. 13, No. 2, pp. 169-185.
- 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.
- Harzing, A.W.K. (2004) 'Ideal jobs and international student mobility in the enlarged European Union', *European Management Journal*, vol. 22, no. 6, pp. 693-703.
- 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. & 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) *Culture's Consequences. International Differences in Work-Related Values*, London: SAGE Publications.
- Hofstede, G. (2001) *Culture's 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.
- 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.
- 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.
- Ralston, D.A., Cunniff, M.K. & 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.
- Ronen, S.; 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. & 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.
- 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 superior's 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.