Original Article

Cultural and age-related differences in reliability: An empirical study in the United States, Japan and Malaysia

Received (in revised form): 5th October 2011

George P. Moschis

is Alfred Bernhardt Research Professor of Marketing and Director of the Center for Mature Consumer Studies, Georgia State University, USA.

Fon Sim Ong

is Professor of Marketing at the Graduate School of Business, Universiti Tun Abdul Razak. She is also an Honourary Research Fellow with the Asia Pacific Institute of Ageing Studies (APIAS), Lingnan University, Hong Kong. She is also Research Fellow with the Graduate School of Business, Universiti Tun Abdul Razak. Her major research interest is in older consumers. She has published numerous articles in journals, conference proceedings, and chapters of book.

Masoud Abessi

is an Associate Professor in the Faculty of Industrial Engineering at Yazd University, Yazd, Iran. His research interests are data mining, MIS and data analysis. He has taught as a visiting professor in several universities, promoting interdisciplinary research and interchangeability applications of quantitative techniques in different fields such as marketing, finance, strategy and MIS.

Takako Yamashita

is an Associate Professor of Marketing in the Faculty of Commerce at the University of Marketing and Distribution Science, Japan.

Anil Mathur

is Brodlieb Distinguished Professor of Business at the Frank G. Zarb School of Business, Hofstra University. He has published more than 75 articles in journals and papers in conference proceedings. Dr Mathur is also the author of numerous research monographs, industry reports, and proprietary consulting reports. He has won many awards for his research and teaching.

ABSTRACT The cross-cultural applicability of measures has received limited attention, though it represents a source of bias in research findings. Although several assumptions and speculations have been made regarding the sources of potential bias in measures used when studying cultural and subcultural differences, little empirical work supports these views. The present study is empirically motivated. It focuses on testing hypotheses derived from theories presented by leading scholars regarding possible sources of differences in reliability coefficients, using cross-sectional data from three countries (the United States, Japan and Malaysia) to falsify them. The tests of the hypotheses regarding cultural, contextual and aging effects provide some results in line with the hypotheses, but they raise additional concerns about the possible sources of measurement differences and the applicability of measures developed in the United States to other countries and subcultures within countries.

Journal of Targeting, Measurement and Analysis for Marketing (2011) **19**, 141–151. doi:10.1057/jt.2011.15

Keywords: reliability; measurement; cultures; subcultures; aging

Correspondence: Fon Sim Ong

Marketing Department, Taylor's Business School, Taylor's University,
1 Jalan Taylor's, Subang Jaya, Selangor 47500, Malaysia
E-mail: fonsim.ong@taylors.edu.my

INTRODUCTION

Although interest in cross-cultural research has increased substantially in recent years, the cross-cultural applicability of the measures employed has received limited attention and represents a source of bias in research findings (for example, Schwarz,¹ Steenkamp and Baumgartner² and Wong *et al*³). Several widely used measures that contain mixed-worded items (MWI) – that is, positively worded items (PWI) and reverse-worded items (RWI), such as materialism and well-being – seem to work well among US respondents, but cross-cultural consumer researchers and psychologists find these scales deficient in other settings, demonstrating questionable reliability and validity (for example, Wong *et al*,³ Eastman,⁴ Ger and Belk,⁵ Iwata *et al*⁶ and Nakano⁷). For example, recent studies show that MWI measures of psychological functioning (for example, optimism, anxiety) that are unidimensional for US respondents split into separate positive and negative factors for Asian Americans (for example, Wong *et al*,³ Nakano,⁷ Cheng and Hamid⁸ and Suzuki *et al*⁹).

Research by Wong *et al*³ addresses issues in using mixed-worded scales in cross-cultural research, as well as the need for additional research. These and other researchers (for example, Schwarz¹) also suggest that problems of MWI scales most often arise for members of subcultural groups such as ethnic, age and racial minorities. Although prior research suggests that problems associated with MWI scales are more pronounced in certain cross-cultural and subcultural settings, the reasons these occur are unclear.³ Several explanations for why MWI formats might create problems for certain groups of consumers have been suggested; they include differences in the difficulty in understanding, interpreting and responding to RWI (for example, Suzuki *et al*⁹ and Javeline¹⁰), possibly as a result of age, cultural differences, religious beliefs or contextual factors (for example, Schwarz¹ and Wong *et al*³). However, though such assumptions seem intuitively appealing, they have been less than rigorously tested.

The present study is empirically motivated. Rather than assessing various theoretical

explanations for differences in responses, it provides data aimed at testing hypotheses based on theoretical grounds proposed by leading researchers and helps answer practical questions such as: Do mixed-item scales produce lower reliabilities in collectivistic countries, where consumers tend to show higher acquiescence, in comparison with consumers in individualistic countries? Do such scales produce lower reliabilities when administered to older adults, who are likely to experience cognitive declines? Are there cultural and age differences in reliabilities owing to question context?

Our study addresses these cultural and age-related issues by examining responses to measurement instruments by consumers of different age-groups in three countries: the United States, Japan and Malaysia. First, we examine the reliability of the measures of self-esteem, life satisfaction and materialism. In the 10-item self-esteem (SSE) scale, half of the statements are reverse-worded; it was developed by Rosenberg¹¹ and has been widely used by social scientists. The scale for life satisfaction is Diener *et al*'s¹² satisfaction with life scale (SWLS), which has been used in studies that have appeared in leading journals (for example, Wong *et al*³) and has demonstrated strong psychometric properties.¹³ Moreover, SWLS has demonstrated adequate cross-cultural conceptual equivalence, holding similar meanings across cultures.^{3,14,15} The third scale is Richins and Dawson's¹⁶ scale of materialism values (MVS), changed into an interrogative question format as developed and used by Wong *et al*.³

HYPOTHESES

This section provides background information on possible sources of bias in measures suggested by the work of previous researchers. Although we recognized that there might be a variety of causes for the observed differences in responses to measures, the theoretical assessment of the various views is beyond the scope of the present study. Instead, the present study is empirically motivated and tests select theories and propositions advanced by leading scholars in different cultural contexts.

Cultural values

One potential source of bias in measurement has been attributed to cultural differences in responding to instruments. For example, Wong *et al.*³ find that East Asians interpret (or respond to) PWI and RWI of Richins and Dawson's¹⁶ scale of materialism differently than do US respondents. In parts of Asia, appearing polite is expected, and agreeableness is an important social norm. Researchers believe that this social norm may prompt subjects to agree with both PWI and RWI.^{17,10} This tendency, which is believed to be present not only among Asian but also in Latin American consumers who have collectivistic values (for example, Marin *et al.*¹⁸), is believed to account for the results reported by Wong *et al.*³ who found that replacing both PWI and RWI scales with a set of questions stated in an interrogative format enhances the validity of MWI scales in cross-cultural settings.

On the basis of these assertions and results of previous studies, we expect our mixed-worded measure of SSE to show cultural differences in the form of varied reliabilities between (i) the United States and Malaysia and (ii) the United States and Japan, because people in East Asian countries reportedly have the tendency to agree with both RWI and PWI. Thus, for the SSE measure, we expect a higher reliability for the US sample than for the remaining two countries.

Hypothesis 1 (H1): The reliability coefficient of the SSE scale is higher for the US sample than for the (a) Malaysian sample and (b) Japanese sample.

Contextual factors

Contextual factors may be sources of bias in measurement. In particular, Schwarz¹ asserts that question redundancy, by which an earlier question inquires about a subject of information (for example, satisfaction with one's job) that is relevant to a subsequent, more general one (for example, life satisfaction), may produce different effects in East Asian versus Western cultures. These differences in question redundancy have been attributed to cultural values. Specifically, it has been suggested that for the individualistic US

culture, which values individual achievement, question redundancy that presents MVS before SWLS is likely to produce a stronger correlation than when the order of questions is reversed. However, in the collectivist cultures of East Asia, this order of presentation should not affect the size of the correlations between SWLS and MVS.¹ If cultural values account for such differences we can expect question redundancy to be a source of bias in the individualistic US culture.

Hypothesis 2 (H2): The size of the negative correlation between MVS and SWLS differs by order of presentation only for the US sample, with the correlation being greater when the MVS scale is presented before rather than after the SWLS scale.

Aging effects

The rapidly growing number of older consumers globally and the increasing need to understand consumer behavior across cultures and age subcultures have resulted in a growing number of studies that address age-related differences in cognitive functioning, decision making and other consumer behaviors (for a review see Schwarz¹). However, many of these age-related differences may be results of the methods used. For example, Schwarz¹ reviews studies that indicate that some observed differences between younger and older age-groups may be due to the instrument, such as question-order effect and frequency estimation. Furthermore, because human aging is accompanied by profound cognitive changes, such as decreased sensory functioning and decline in working memory capacity (for a review, see Park¹⁹), age differences reflect performance on each of the tasks involved in answering questions, such as question comprehension, recall of relevant information, judgment formation and response editing.²⁰ However, two areas of potential age-related bias in surveys that have received less than adequate attention are the use of RWI scales and question-order effects.

Responses to RWI. Researchers who employ MWI often find that RWI display somewhat lower reliability and, when subjected to factor

analysis, the RWI load on a separate factor.³ Although MWI scales have appeared problematic for members of subcultural groups, including ethnic and racial minorities, age-related differences in responses have not been investigated.¹ Because cognitive aging appears to be a universal process, our main contention is that bias in MWI scales is present to a greater extent among older than among younger age-groups regardless of cultural context. Schwarz¹ theorizes that this increase in bias with age may be due to the mixed-item-order effect, because buffering PWI with RWI requires the older respondent to change his or her judgment and response process to consider multiple alternatives, a more demanding task for older adults who have experienced declines in working memory. Wong *et al*³ further suggest that if RWI present greater difficulty than PWI, in general, RWI should display substantially weaker within-group differences than PWI. Specifically, on the bases of these views, we expect the differences between RWI and PWI in the within-group correlations to be weaker among older than among younger age-groups. Thus, we expect MWI scales to show a lower internal consistency. According to this expectation, our MWI scale of SSE should demonstrate a lower internal consistency than the SWLS scale, although the former scale contains twice as many items. Age-related differences in responses to SWLS, which contains five PWI, should not be significantly different between the two age-groups. In addition, the differences in the within-group negative correlations between PWI and RWI of the SSE scale should be stronger among younger than among older consumers.

Furthermore, because scales containing items that have an interrogative response format increase the internal consistency of a scale that initially contained MWI,³ we expect the internal consistency of the SSE scale, with its equal number of RWI and PWI, to be lower among older than among younger consumers. The difference in the internal consistency of MVS may be smaller between the two age-groups than the internal consistency differences for the SSE

scale. The preceding discussion suggests the following hypotheses:

Hypothesis 3 (H3): The negative correlations between RWI and PWI of the SSE scale are stronger among the younger samples than among the older samples of consumers in (a) the United States, (b) Malaysia and (c) Japan.

Hypothesis 4 (H4): Age differences in internal consistency are greater for the SSE scale than for MVS, with younger age-groups showing higher reliabilities than older age-groups of consumers in (a) the United States, (b) Malaysia and (c) Japan.

Question-order effects

Question-order effects suggest bias in responses when preceding questions remind the respondent of information that otherwise he or she would not consider in answering a subsequent question.²¹ The available data support the assumption that age-related declines in memory attenuate question-order effects.¹ However, we contend that question-order effects are contingent on the subject to which questions pertain. With respect to the influence of the materialism-life satisfaction question format, for example, we expect that because older US adults are likely to be less materialistic than their younger counterparts (for example, Burroughs and Rindfleisch²²), the influence of the MVS-SWLS order will be more significant among younger than among older US adults. In contrast, to the extent consumers in Eastern cultures, in general, are less materialistic,^{23,3} there should be no significant difference in the correlations between the two scales among older versus younger Malaysian and Japanese respondents; moreover, the influence of the order of presentation should be greater among US consumers than among their Malaysian and Japanese counterparts, regardless of age.

Hypothesis 5 (H5): When MVS is presented before SWLS, the correlation between the two scales is stronger for younger than for older samples of US consumers.

Hypothesis 6 (H6): Regardless of age, when MVS is presented before SWLS the correlations between the two scales is stronger for the US sample than for the (a) Malaysian and (b) Japanese samples.

METHODS

We chose two rather diverse countries, Malaysia and Japan, to represent the collectivist countries of Asia that differ markedly with respect to ethnicity and religion. Japan is a relatively closed and homogeneous culture, whereas Malaysia is a diverse culture comprising several subcultures – mainly Malays, Chinese and Indians. With respect to religious beliefs, Japanese are homogeneous, while Malaysians again are diverse (Muslims, Hindus, Buddhists and other religions). As Asian countries, both of them are characterized as having collectivist values.^{24–26} For our research objectives, cross-sectional surveys appear acceptable²⁷ and so is the use of multiple methods of data gathering (for example, Wong *et al.*³). Although method and sample equivalence, as suggested in cross-cultural research, is a desirable goal, we had to compromise our ambitions when faced with the realities and challenges that have confronted previous researchers.³

Samples

We employed convenience sampling and the same method of survey administration in both Asian countries (personal interviews by trained interviewers who attended training sessions to ensure that they understand the questionnaire and method of personal interview). For the survey in Malaysia, we used personal interviews in which respondents were drawn from five major towns: Penang in the north, Kuala Lumpur/Petaling Jaya in the central region, Johor Baru in the south and Kuantan and Kota Baru in the eastern corridor of Peninsular Malaysia. These urban areas are representative of Malaysian urban population as the western parts of Malaysia are developed and the eastern corridor is less developed. The study in Japan adopted the same method of data collection in which convenience sampling was drawn using personal interviews conducted in areas in and around Tokyo. As

Japan is a developed country, data collection in and around Tokyo was considered appropriate. The sample sizes were as follows: US ($N=323$), Malaysia ($N=539$: Malays = 261, Chinese = 204, Indians = 74), and Japan ($N=728$). The Malaysian questionnaires were available in both English and Malay languages. Measurement scales used in the Japanese questionnaires included those available in translated versions, which have been statistically confirmed as valid.^{28–30} For scales that had no formal Japanese translation, we used back-translation methods.

Measures

We used MVS that was developed and used by Wong *et al.*³ who survey consumers from several Asian countries and find that replacing items posed as statements with items framed as questions enhances the validity and reliability of this scale. We used the same 15 items used by Wong *et al.*,³ as well as identical instructions and response formats for these items. The alpha reliability coefficients of this scale for the three countries are: Japan = 0.669, Malaysia = 0.838 and US = 0.822. For Rosenberg's¹¹ SSE, the reliability coefficients are 0.668, 0.888 and 0.864 for the three countries, respectively; we used Diener *et al.*'s¹² SWLS with alpha reliability coefficients of 0.872, 0.859, and 0.906, respectively.

Because of respondent sensitivity to 10 items of the SSE scale, especially in personal interviews with our Asian respondents, we used the short form of the Marlowe–Crowne social desirability scale³¹ in the surveys administered to the Eastern samples of respondents. This social desirability scale, which contains items of high sensitivity and appears preferred among consumer researchers (for example, Middleton and Jones³²), also consists of 10 items, 5 scored as 'true' and 5 as 'false'. We changed the response format of this scale to the same four-point 'agree-disagree' scale as SSE, and we used it as a validity check for the SSE scale. The alpha reliability coefficient of this scale is 0.686 for the sample from Japan and 0.684 for the sample from Malaysia. These alpha reliability coefficients are somewhat lower than values obtained from Western samples and are consistent with the findings of previous studies.³²

Thus, MWI of SSE appears to present no significant problems for Malaysia, but the same response format of the social desirability scale does. In a similar vein, while the interrogative question format appears to produce acceptable reliability coefficients for Malaysia and the United States, it does not for Japan.

RESULTS AND ANALYSES

Before hypothesis testing, we computed the basic statistics and correlations for the variables used in our hypotheses for each of the three countries. Tables 1, 2 and 3 show the means and standard deviations for the variables used in the hypotheses for the US, Malaysian and Japanese samples, respectively. We assumed equality of variances of our measures across the three countries because of absence of any strong test of variance equality, owing to insensitivity of correlation and regression analyses to the equality of variances.^{33,34}

To test the correlations between PWI and RWI for Hypothesis 3, we adopted a structural equation modeling (SEM) approach. Two constructs were developed, one for PWI of the SSE scale and the other for RWI. Fit indices ensure good fits and proper representations of the population under study by the samples collected. Compared with the alternative and conventional methods used to compute the correlations between such constructs, which employ factor scores derived from either score summation or regression weights after performing exploratory factor analysis, SEM instead corrects for the measurement errors and provides more accurate

correlation estimates. In short, less error gets introduced in the correlation computation process by using the SEM approach than the conventional approach reported in the literature.^{35,36}

Table 4 provides a summary of the hypotheses of this study and the results. Hypothesis 1 posits that the reliability coefficient of the SSE scale is

Table 1: Descriptive statistics for the US sample

Variable	Total		Young		Old	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
SSE	38.07	5.4	37.90	6.0	38.20	5.29
PWI	16.73	2.62	16.52	2.96	16.83	2.48
RWI	8.66	3.50	8.62	3.46	8.64	3.50
MVS	39.12	8.95	41.12	9.16	38.31	8.74
SWLS	17.49	5.04	16.45	5.17	17.85	4.98
AGE	61.3	12.8	46.1	7.9	66.7	8.9
N	313		84		226	

Abbreviations: SSE – self-esteem scale; PWI – positively worded items; RWI – reverse-worded items; MVS – material values scale; SWLS – satisfaction with life scale.

Table 2: Descriptive statistics for the Japanese sample

Variable	Total		Young		Old	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
SSE	28.84	3.46	28.26	3.48	29.47	3.34
PWI	14.68	1.66	14.52	1.66	14.86	1.65
RWI	14.36	1.69	14.29	1.61	14.44	1.77
MVS	45.16	6.87	45.88	7.13	44.52	6.48
SWLS	16.96	5.32	14.47	3.97	19.56	5.96
AGE	52.1	15.5	39.3	9.2	65.4	7.1
N	723		366		352	

Abbreviations: SSE – self-esteem scale; PWI – positively worded items; RWI – reverse-worded items; MVS – material values scale; SWLS – satisfaction with life scale.

Table 3: Descriptive statistics for the Malaysian sample

Variable	Total		Young		Old		Buddhists		Muslims		Non-Buddhists	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
SSE	31.02	5.36	31.24	5.12	30.55	5.79	29.13	4.07	32.82	5.68	31.89	5.65
PWI	16.26	2.62	16.35	2.45	16.06	2.94	15.27	2.14	17.15	2.66	16.71	2.70
RWI	14.76	3.74	14.89	3.20	14.49	3.34	13.86	2.57	15.67	3.44	15.18	3.44
MVS	46.83	6.26	47.92	6.21	44.60	5.75	48.07	6.59	46.13	5.58	46.25	6.01
SWLS	23.05	5.96	22.43	6.07	24.32	5.54	21.23	6.07	24.74	5.62	23.89	5.72
AGE	47.0	15.4	38.1	9.1	65.2	7.6	47.7	15.9	47.5	15.5	46.6	15.1
N	539		362		177		171		261		368	

Abbreviations: SSE – self-esteem scale; PWI – positively worded items; RWI – reverse-worded items; MVS – material values scale; SWLS – satisfaction with life scale.

Table 4: Summary of hypotheses, results and hypotheses confirmation (Yes) or non-confirmation (No)

	Expectations	Results			Tests and significance level		
H1:	Alpha of SSE is higher for US sample than for (a) Malaysians (MY) and (b) Japanese (JN)	(a) US (0.864) versus MY (0.888) (b) US (0.864) versus JN (0.668)			$F=1.214$ $F=2.441$	0.036 0.001	No Yes
H2:	When MVS is presented before SWLS, the correlation is stronger than when the scales are presented in a reverse sequence among US consumers	MVS first $r=-0.03$ SWLS first $r=-0.22$			$Z=1.713$	0.043	No
H3:	The negative correlation between PWI and RWI of SSE is stronger among younger than among older consumers in (a) the United States, (b) Malaysia and (c) Japan	(a) US	Young $r=-0.37$ Old $r=-0.33$		$Z=-0.343$	0.363	No
		(b) Malaysia	Young $r=-0.66$ Old $r=-0.77$		$Z=-2.463$	0.007	No
		(c) Japan	Young $r=-0.21$ Old $r=-0.57$		$Z=-5.056$	0.001	No
H4:	Regardless of country, age difference in alphas are present for SSE but not for MVS, with younger age-groups showing higher reliabilities than older age-groups in (a) US, (b) Malaysia and (c) Japan	(a) US – SSE	Y=0.747	O=0.726	$F=1.08$	ns	No
		US – MVS	Y=0.833	O=0.815	$F=1.10$	ns	Yes
		(b) Malaysia – SSE	Y=0.874	O=0.910	$F=1.40$	0.006	Yes
		Malaysia – MVS	Y=0.819	O=0.844	$F=1.16$	ns	Yes
		(c) Japan – SSE	Y=0.840	O=0.831	$F=1.05$	ns	No
		Japan – MVS	Y=0.676	O=0.611	$F=1.20$	0.04	No
H5:	When MVS is presented before SWLS, the correlation between the two scales is stronger among younger than among older US respondents	US $Y=-0.263$	O=-0.220		$Z=-0.252$	0.390	No
H6:	Regardless of age, when MVS is presented before SWLS the correlation between the two scales is stronger for the US sample than for the (a) Malaysian and (b) Japanese samples	Y: (a) US versus MY=-0.263	versus -0.212		$Z=-0.317$	ns	No
		(b) US versus JN=-0.263	versus -0.169		$Z=-0.582$	0.27	No
		O: (a) US versus MY=-0.220	versus -0.530		$Z=2.514$	0.006	No
		(b) US versus JN=-0.220	versus -0.078		$Z=-1.174$	0.120	No

Note: Y=Young, O=Old.

greater for the US sample than for the (a) Malaysian and (b) Japanese samples. We tested for differences in the alpha reliability coefficient of SSE for the US sample versus the (a) Malaysian and (b) Japanese samples. Contrary to our expectations, the alpha coefficient for the US sample is somewhat smaller (0.864) than that for the Malaysian sample (0.888), yielding a Feldt's F -value of 1.21 for the difference between the two coefficients that is statistically significant ($P<0.036$). In contrast, the alpha coefficient of the SSE scale for the Japanese sample is lower (0.668), as we expected, producing a Feldt's F -value of 2.44 for the differences between the two coefficients, which is statistically significant ($P<0.001$). Thus, Hypothesis 1(a) is not supported, whereas Hypothesis 1(b) is supported.

Hypothesis 2 posits that the size of correlations between MVS and SWLS differs by order of

presentation only for the US sample, with the correlation being greater when the MVS scale appears before rather than after the SWLS scale. We tested for differences in the relationships (correlations) between the MVS and SWLS scales for two subgroups of respondents *separately* (that is, those who answered MVS first followed by SWLS versus those who responded to the same questions in the reverse order). We expected the group that answered the MVS questions first to show a stronger correlation between the two scales than the group that answered the questions for these scales in reverse order in the US subgroups but not in the same subgroups of the other two countries. However, the data do not provide support for this hypothesis; a correlation of -0.031 emerges between the two constructs when MVS appears first, and a correlation of -0.22 results when SWLS is presented first

($Z=1.7212$, $P<0.043$). The group of US respondents that answered the MVS questions first before answering the SWLS questions does not show a stronger correlation, and therefore the hypothesis is not supported.

Hypothesis 3 pertains to the relationship between PWI and RWI among younger and older consumers in the three countries. The negative correlations between RWI and PWI of the SSE scale should be stronger among the younger samples than among the older samples of the three countries, but according to the results shown in Table 4, the difference in the correlations of the two age-groups is not statistically significant for the US sample. For the Malaysian and Japanese samples, though the difference is statistically significant, the results move in the opposite direction than we expected (stronger for older adults). Thus, Hypothesis 3 is not supported.

Hypothesis 4 focuses on the differences in the reliability coefficients between two scales (SSE and MVS) and predicts age differences in SSE but not in MVS for each of the three countries. For the Malaysian sample, the results show age difference in alphas for the SSE scale but not for MVS (Feldt's $F=1.40$, $P<0.006$; $F=1.16$, ns), respectively, in support of the hypothesis. However, for the Japanese sample, the results indicate significance for MVS but not for SSE (contrary to the predicted direction), providing no support for the hypothesis. For the US sample, the age difference in alphas is not significant for the SSE scale (Feldt's $F=1.08$, ns); for the MVS scale, the results reveal no age difference, as expected (Feldt's $F=1.10$, ns). Thus, Hypothesis 3(b) is supported, but Hypotheses 3(a) and 3(c) are not.

Hypothesis 5 states that when MVS appears before SWLS, the difference in the correlations between the two scales will be greater for younger than older samples of US consumers. That is, the younger group of US respondents who answer the MVS questions before answering the SWLS questions should show a stronger correlation between the two scales than the older group that answers the questions *in the same order*. We tested this hypothesis by comparing the

resultant correlation between the two scales in the older sample (age 55 years and over) ($r=-0.263$) with the correlation in the younger sample ($r=-0.220$). This comparison produces a Z score of -0.252 (ns) and thus no support for Hypothesis 5.

The last hypothesis posits that, regardless of age, when the MVS is presented before the SWLS scale, the difference in the correlations between the two scales is greater for the US sample than for the Malaysian or Japanese samples. We first computed correlations for the two scales for the US, Malaysian and Japanese respondents, separately for both younger and older subsamples. Next, we performed a series of Z -transformations: US versus Malaysia and US versus Japanese. The results of these comparisons are shown in Table 4. The Z -transformations show that the US sample does not have a stronger correlation than the Malaysian sample for the younger age-group ($Z=-0.317$, ns), providing no support for Hypothesis 6(a). However, for the older age-group, we find a stronger correlation between the two scales ($Z=2.51$, $P<0.006$), though not in the predicted direction. Similarly, comparisons of the correlations between the US and Japanese samples fail to support Hypothesis 6(b), because those correlations are not stronger for the US sample than for the Japanese sample ($Z=-0.582$, ns; $Z=-1.17$, ns, for the young and old age-groups, respectively). Thus, Hypotheses 6(a) and 6(b) are not supported.

DISCUSSION

The cross-cultural applicability of measures represents a challenge for researchers. Although several assumptions and speculations suggest sources of potential bias in the measures used when studying cultural and subcultural differences, little empirical work has been published in support of these views. The present study is an empirically motivated effort to validate several assumptions about various sources of bias. Our research tests several hypotheses that are either based on propositions by previous researchers or derived from analyses and interpretations of previously published research. They focus primarily on the reliability of three

measures developed by US researchers: SSE, which employs an equal number of PWI and RWI; SWLS, which consists of only PWI; and MVS, which uses an interrogative question format rather than mixed-wording items.

We expected the alpha reliability coefficient of the SSE scale to be higher for the US sample than for the Malaysian and Japanese samples, but our data show that this difference occurs only for the Japanese sample. Specifically, the mixed-wording scale of SSE has a higher (rather than the expected lower) reliability compared with the MVS and SWLS scales among Malaysian consumers. But SSE is less reliable than SWLS (but not MVS) among the Japanese, as expected.

We also examine the contextual effects of question ordering by comparing the correlations between MVS and SWLS. The relationship between MVS and SWLS differs across countries (both in direction and magnitude of influence), yet nearly every study conducted in the United States has found a negative relationship between these variables. Thus, we expected a negative relationship between MVS and SWLS among consumers in the US sample. Furthermore, we expected the strength of the negative association to differ according to the order of presentation, with a stronger correlation when MVS appears before SWLS. Our US data instead show the negative relationship between the two variables only when MVS is presented *after* SWLS.

We examine the effects of age on consumer responses, expecting several age-related differences, including that the negative correlations between PWI and RWI of the SSE scale would be stronger among younger consumers than among older consumers, regardless of country. Our expectations attain no support from any one of our samples. For Malaysia and Japan, the older samples show a stronger correlation between the two subsets of the SSE scale compared with the younger subsamples. We compare the alphas of the SSE scale, which contains an equal number of PWI and RWI, with MVS, which contains modified items to avoid the problem of social desirability, expecting significant differences in the reliabilities of the SSE scale (but not in the alphas of MVS) across the three countries.

However, the alpha coefficients of the SSE scale and the MVS scale do not differ significantly between the two age subsamples of the three countries. In fact, for the Malaysian sample, SSE shows a higher alpha among the older than the younger sample.

Finally, we examine the effects of age on the order of presentation of questions designed to measure life satisfaction and materialism. We expected the correlations between the two measures (SWLS and MVS) to be stronger among younger than among older consumers when MVS appears before SWLS for the US sample, but our data do not support this hypothesis. We also test the effects of age on the same order of presentation, expecting a stronger negative correlation among US consumers than among consumers of Malaysia and Japan. Again, our data do not support these expectations.

Our study is not without limitations. Despite our efforts to apply procedures that reduce error in our derivation of the correlation coefficients, our findings may have been influenced by errors owing to cultural factors in the data collection process and the measurement instruments. Although the data collection methods (personal interviews for Malaysia and Japan, self-administered for the United States) may create a bias, that influence would have affected only Hypothesis 1, which compares the US sample with the samples from each of the Asian countries. Yet, to the extent that cultural differences tend to produce *systematic errors* – that is, consistency in response tendency to different methods and response formats (for example, Asian respondents tend to show response biases such as acquiescence and extreme responding – for example, Baumgartner and Jan-Benedict³⁷), they may not have affected reliabilities (p. 387).³⁸ In this case, any bias should have been consistent in comparing the alphas of SSE between US respondents versus Malaysian and Japanese respondents. However, this consistency was not the case; this hypothesis received support for the Japan versus US comparison only. The sizes of the alpha coefficients for the US versus Malaysia comparison were opposite to our expectations. It might be that consumers in different cultures

interpreted the items of this scale differently, or this measure may not adequately capture self-esteem across countries (for example, Crocker and Park³⁹). In a similar vein, although the self-administered method of data gathering for the US sample may create bias that would have affected the results concerning the effects of age on the order of presentation, we had no reason to believe that either age-group is inclined to answer self-administered survey questions in reverse sequence. It is possible that different age-groups may have interpreted and responded to scale items differently.

Further research might help determine the applicability of the data collection methods and instruments that we used to study consumers in different cultures and countries, such as question format, response format and question redundancy. Such information would help us develop better instruments to study consumers in different age-groups, countries (cultures) and subcultures. For those hypotheses that were confirmed, more rigorous testing would be needed and alternative explanations for the confirmed relationships should be considered, because cross-sectional data, such as ours, can only falsify hypothesized relationships rather than proving them.⁴⁰ Interpretive research could shed light on how consumers in different cultures respond to items of scales such as SSE and how different age-groups interpret the meaning of self-esteem items. Such research could lead us to the development of revised or different scale to measure self-esteem in collectivistic countries of Asia.

In summary, the study findings highlight the challenges that researchers confront when conducting cross-cultural studies. The sample of countries, subcultures and instruments used in this research may be limited, but our findings raise additional concerns about the applicability of measures developed in one country to other countries and subcultures within countries. Assertions regarding the influence of age and cultural and subcultural factors on consumer responses to various instruments appear intuitively appealing, and some are even suggested by theory and research, but many such assumptions must be

rigorously tested in different settings before these instruments can gain acceptance as tools for further cross-cultural research.

REFERENCES

- Schwarz, N. (2003) Self-reports in consumer research: The challenge of comparing cohorts and cultures. *Journal of Consumer Research* 29(March): 588–594.
- Steenkamp, J. and Baumgartner, H. (1998) Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research* 25(1): 78–90.
- Wong, N., Rindfleisch, A. and Burroughs, J. (2003) Do reverse-worded items confound measures in cross-cultural consumer research? The case of material value scale. *Journal of Consumer Research* 30(June): 72–91.
- Eastman, J.K., Fredenberger, B., Campbell, D. and Calvert, S. (1997) The relationship between status consumption & materialism: A cross-cultural comparison of Chinese, Mexican & American students. *Journal of Marketing Theory & Practice* 5(Winter): 52–66.
- Ger, G. and Belk, R.W. (1996) Cross-cultural differences in materialism. *Journal of Economic Psychology* 17(February): 55–77.
- Iwata, N., Roberts, C.R. and Kawakami, N. (1995) Japan-US comparison of responses to depression scale items among adult workers. *Psychiatry Research* 58(October): 237–245.
- Nakano, K. (2001) Psychometric evaluation of the Japanese adaptation of the aggression questionnaire. *Behavior Research and Therapy* 39(July): 853–858.
- Cheng, S.-T. and Hamid, P.N. (1997) Dispositional optimism in Chinese people: What does the life orientation test measure? *International Journal of Psychology* 32(1): 15–22.
- Suzuki, T., Tsukamoto, K. and Kazuhiro, A. (2000) Characteristic factor structures of the Japanese version of the State-Trait Anxiety Inventory: Coexistence of positive-negative and state-trait factor structures. *Journal of Personality Assessment* 74(3): 447–458.
- Javeline, D. (1999) Response effects in polite cultures: A test of acquiescence in Kazakhstan. *Public Opinion Quarterly* 63(1): 1–28.
- Rosenberg, M. (1965) *Society and Adolescent Self-image*. Princeton, NJ: Princeton University Press.
- Diener, E., Emmons, R., Larsen, R. and Griffin, S. (1985) The satisfaction with life scale. *Journal of Personality Assessment* 49(1): 71–75.
- Robinson, J., Phillip, P., Shaver, R. and Wrightman, L.S. (eds.) (1991) Measures of personality and social psychological attitudes. In: *Measures of Social Psychological Attitudes*, Vol. 1. San Diego, CA: Academic Press.
- Oishi, S., Schimmack, U., Diener, E. and Suh, E.M. (1999) The measurement of values and individualism-collectivism. *Personality and Social Psychology Bulletin* 24(November): 1177–1189.
- Suh, E.M., Diener, E., Oishi, S. and Triandis, H.C. (1998) The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology* 74(February): 482–493.
- Richins, M.L. and Dawson, S. (1992) A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of Consumer Research* 19(December): 303–316.

- 17 Hui, C.H. and Triandis, H.C. (1983) Multistrategy approach to cross-cultural research: The case of locus of control. *Journal of Cross-Cultural Psychology* 14(March): 65–83.
- 18 Marin, G., Gamba, R.S. 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.
- 19 Park, D. (2000) The basic mechanisms accounting for age-related decline in cognitive function. In: D. Park and N. Schwarz (eds.) *Cognitive Aging: A Primer*. Philadelphia, PA: Psychology, pp. 3–22.
- 20 Schwarz, N. and Knauper, B. (2000) Cognition, aging and self-reports. In: D. Park and N. Schwarz (eds.) *Cognitive Aging: A Primer*. Philadelphia, PA: Psychology, pp. 233–252.
- 21 Sudman, S., Bradburn, N.M. and Schwarz, N. (1996) *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass.
- 22 Burroughs, J.E. and Rindfleisch, A. (2002) Materialism and well-being: A conflicting values perspective. *Journal of Consumer Research* 29(December): 348–370.
- 23 LaBarbera, P.A. and Gurhan, Z. (1997) The role of materialism, religiosity and demographics in subjective well-being. *Psychology and Marketing* 14(1): 71–97.
- 24 Huff, L. and Kelley, L. (2003) Levels of organizational trust in individualist versus collectivist societies: A seven-nation study. *Organization Science* 14(1): 81–90.
- 25 Kacen, J.J. and Lee, J.A. (2002) The influence of culture on consumer impulse buying behavior. *Journal of Consumer Psychology* 12(2): 163–176.
- 26 Ngai, E.W.T., Heung, V.C.S., Wong, Y.H. and Chan, F.K.Y. (2007) Consumer complaint behavior of Asians and Non-Asians about hotel services: An empirical analysis. *European Journal of Marketing* 41(11/12): 1375–1391.
- 27 Rindfleisch, A., Malter, A.J., Ganesan, S. and Moorman, C. (2008) Cross-sectional versus longitudinal survey research: concepts, findings, and guidelines. *Journal of Marketing Research* 45(August): 261–279.
- 28 Diener, E., Suh, E. and Oishi, S. (1997) Recent findings on subjective well-being. *Indian Journal of Clinical Psychology* 24(March): 25–41.
- 29 Kitamura, T. and Suzuki, C. (1986) Nihongo ban social desirability scale ni suite. *Syakai Seishi Igaku* 9: 173–180.
- 30 Nakagawa, S. (1990) Jiko keisei no shinrigaku. *TASC Monthly* May: 4–11.
- 31 Strahan, R. and Gerbasi, K.R. (1972) Short, homogeneous version of the Marlowe–Crowne Social Desirability Scale. *Journal of Clinical Psychology* 28(2): 191–193.
- 32 Middleton, K.L. and Jones, J.L. (2000) Socially desirable response sets: The impact of country culture. *Psychology and Marketing* 12(2): 149–163.
- 33 Bobko, P. and Russell, C. (1990) Variance homogeneity in interactive regression: A clarifying note about data transformations. *Journal of Applied Psychology* 75(February): 95–96.
- 34 Cohen, J. and Cohen, P. (1983) *Applied Multiple Regression/correlation Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.
- 35 Alsawalmeh, Y.M. and Feldt, L.S. (1999) Testing the equality of two independent α coefficients adjusted by the Spearman–Brown Formula. *Applied Psychological Measurement* 23(December): 363–370.
- 36 Hair, J.F., Black, B., Babin, B., Anderson, R.E. and Tatham, R.L. (2006) *Multivariate Data Analysis*, 6th edn. New York: Prentice Hall.
- 37 Baumgartner, H. and Jan-Benedict, E.M.S. (2001) Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research* 38(May): 143–156.
- 38 Churchill, G.A. (1987) *Marketing Research*. Chicago, IL: The Dryden Press.
- 39 Crocker, J. and Park, L.E. (2004) The costly pursuit of self-esteem. *Psychological Bulletin* 130(3): 392–414.
- 40 Popper, K. (1959) *The Logic of Scientific Discovery*. New York: Basic Books.