

CULTURE-BASED SURVEY RESPONSE BIAS: CLARIFYING DIVERGENT FINDINGS IN COMPARISONS OF SAMPLES FROM NORTH AMERICA AND CONFUCIAN ASIA

**Robert A Culpepper, Stephen F Austin State University
Joseph K Ballenger, Stephen F Austin State University**

ABSTRACT

We reviewed nine studies relevant to a neglected but crucial methodological concern in all cross-national research employing survey methodology, culture-related response bias. Specifically, we focused on response biases thought to be in play when comparing survey results of North American respondents and those from Confucian Asia, including China, Japan, Singapore, Taiwan, and South Korea.

Our primary purpose was to examine empirical studies that potentially lend support to either of two conflicting theoretical positions in the literature. One view, based on a cultural trait called Asian overconfidence, holds that subjects from countries with a historical legacy of Confucianism tend to exhibit extreme response bias on Likert-type, variable-response items. For example, given a seven-point scale, respondents would tend to choose “1” s and “7” s. An opposing theoretical position predicts that traditional Confucian modesty norms prevalent in these same countries will lead to a cautious, or midpoint, bias for survey respondents. These two diametrically opposite positions have coexisted in the literature for some time, each taking little to no cognizance of the opposing view. We suggest that these seemingly incompatible positions in fact may both be empirically tenable, although contingent on the type of rating task employed in survey items.

Theory suggests that respondents from countries in Confucian Asia revert to a sense of accepted folk wisdom for value judgments and rather than weighing pros and cons/counterevidence, more common in Western countries. We suggest that this rationale should only apply to survey items requiring agreement or disagreement with propositions that are nomothetic in nature, i.e., having general, if not universal, application. Consistent with this rationale, those studies reviewed in this investigation employing nomothetic-type items displayed extreme response tendency for Asian samples when compared to North American samples.

Other studies reviewed employed idiographic items requiring agreement/disagreement with assertions relating to a specific incident or case, such as one’s self, one’s boss, one’s company and so on, with no general with any general application beyond the specific case in question. For these items, respondents from historically Confucian countries gave more cautious or modest responses than did Americans.

We explain how these findings can help resolve contradictory theory and empirical inferences drawn in the literature. The reviewed studies illustrate how taking differences in item type into account can help resolve the counterintuitive divergence in previous findings related to response bias.

INTRODUCTION

With the voluminous level trade between Western and Pacific Rim nations, cross-cultural research has become increasingly important across a number of disciplines. There has been a particularly strong focus on differences between East Asian and Western management practices. Empirical work in this area must bridge a large cultural chasm that raises a number of methodological difficulties. One of the most formidable is the pervasive use of Likert-type and semantic differential rating scales and their long-recognized, but under-investigated, susceptibility to culture-related response bias (Adler et al., 1989; Jaccard & Wan, 1986; Leung & Bond, 1989; Mullen, 1995; Zax & Takahashi, 1967). The implications of ignoring response tendency differences can be profound, leading to major inferential errors in cross-cultural research (Chun et al., 1974; Cronbach, 1946; Singh, 1995).

One of the most frequently encountered forms of response bias encountered in international research is usually referred to as “*extreme response bias*.” Here, subjects tend to choose extreme points on an item scale, “1” s and “7” s on a seven-point scale, for example. Also commonly discussed and observed is “*midpoint bias*” in which subjects, when presented with a seven-point scale, tend to choose a “4.” Response styles and biases have been studied for several decades in datasets that are not international (Cronbach, 1946; Cronbach, 1950; Hamilton, 1968). What makes response bias a much bigger problem in international studies than in national ones is that a particular response tendency can be systematic across an entire country sample, presumably stemming from cultural traits, and not at all operant in the country samples that are being compared (Chun et al., 1974).

The presence of systematic response bias among respondents from one country and not among respondents from another creates all sorts of trouble, as some have discovered (Adler et al., 1989; Chun et al., 1974; Culpepper et al., 2002; Kotabe et al., 1991; Lee & Green, 1991; Zax & Tanashi, 1967). Results obtained from statistical difference tests (e.g., T-tests or F-tests) as well as all methods based on correlational relationships are highly likely to be spurious (Chun et al., 1974; Singh, 1995; Stening & Everett, 1984; Yu et al., 1993). The use of correlational methods is extremely common, of course, and would include regression, structural equation modeling, exploratory factor analysis, confirmatory factor analysis, and a number of others.

Previous work suggests that study samples from countries with historical Confucian influence, such as China, Korea, Taiwan and Japan are highly subject to survey response bias. Unfortunately, this still nascent body of work is characterized by inconsistent theoretical claims and empirical results. For example, some theory and empirical evidence suggests that respondents from countries with a cultural legacy from historical Confucianism are prone to a midpoint bias stemming from caution and/or Confucian modesty norms when answering ratings to Likert-type survey items (Chen et al., 1995; Hui & Triandis, 1989).

Other work claims to support extreme response bias for Likert-type items, dovetailing with theory and evidence from the decision-making literature addressing “*Asian overconfidence*.” Specifically, Yates, Lee and colleagues Lee et al. (1995); Sieck & Yates (2003); Yates et al. (1996); Yates et al. (1989); Yates et al. (2010) and others Whitcomb et al. (1995); Wright & Phillips (1980) hold that those from cultures in Confucian Asia are overconfident in their judgments and evaluations because these cultures - owing largely to historical Confucian influence Yates & Lee (1996); Yates et al. (1997) - are less adversarial and debate between

opposing viewpoints is seen as undesirable. This rationale holds that decision-makers from countries with a Confucian legacy, rather than constructing and weighing arguments on both sides of an issue, tend to revert to a sense of that which seems time-tested and widely accepted—a sense of “*what everybody knows.*”

In the “*folk knowledge*” view, effects are not just limited to decision-making. An additional major implication is that survey respondents from these countries should be less prone to weigh relevant pros and cons bearing on their response and thus are less prone to choose middle values (e.g., values 3, 4, and 5 on a seven-point scale) on the numerical index, which imply ambivalence or compromise between conflicting rationales or bases of support. Rather, they are thought to revert to a sense of “*what everybody knows*” and choose extreme values on multi-point scales (e.g., 1 and 7 on a seven-point scale), expressing strong dissent or agreement vis-à-vis Likert-type items.

We propose that response bias related to Confucian Asia may depend on the type of rating task employed. Theory describing the matching-to-accepted-wisdom cognitive process, supposedly operant in Confucian cultures, does not seem relevant to items requiring idiographic evaluations of an individual person, unit, group, or company. Such items require judgments about a specific individual or case, judgments that are necessarily more complex and analytical. For example, it is quite common for attitudinal scales to include items exclusively referencing the self, such those found in the typical self-esteem scale—“*I feel that I have a number of good qualities.*”

Idiographic items do not refer to that which has general application and thus do not allow for simple reference to a sense of that which is time-tested or generally accepted. Examples would include items such as “*My boss is an effective manager*”, “*I am assertive*”, “*I am satisfied with my job*”, “*I am highly confident in social situations.*” Thus, items requiring idiographic evaluation should not evoke extreme response tendency—rather, theory predicting more modest and less assertive opinions, stemming from historically Confucian values, would seem relevant to this item type.

In contrast, theory about the Chinese employing a sense of what is generally accepted, known, or time-tested, would seem applicable to items requiring respondents to make some form of nomothetic evaluation. Such items require respondents to rate, for example, the propriety or value of specific rules, principles, or systems, which have general and broad application—for example, “*Established standards should not be questioned*” and “*Return on equity is an effective measure of company performance.*” These can then be handled by simply reverting to a sense of folk wisdom, or “*what everybody knows.*” This should be reflected in exaggerated agreement in the case of Likert-type items and similar extreme responses for semantic differential items, e.g., selecting rating options “*extremely valuable*”, “*extremely useful*”, “*extremely appropriate*”, or “*practiced with extreme frequency.*” Finally, we proposed that extreme response tendency should also apply to matters of specific fact or general knowledge. This also was based on Yates & Lee (1996) overconfidence theory, according to which Chinese subjects, when considering a factual proposition, construct fewer counterarguments and contemplate less contrary evidence to the proposition at hand.

A Look at Relevant Studies Employing Idiographic and Nomothetic Items

For purposes of the current survey, we reviewed studies related to the question of whether the nature of survey items leads to differences in response bias patterns. We were interested in studies that met three criteria. First, we looked for studies that identified survey response bias explicitly- even if it was incidental to the main focus of the study. We also included studies for which the existence of extreme or midpoint response patterns could be reasonably inferred, even when not explicitly addressed by the study's authors. Second, we sought studies which compared U.S. or Canadian samples to samples from historically Confucian countries, i.e., China, Taiwan, Japan, and South Korea. Confining the analysis studies in this this group served to avoid over-generalizing with respect to culture-related patterns that might be evident. Third, we limited studies further to include only those in which it was possible to determine whether items were either nomothetic or idiographic in nature.

We were able to identify ten studies which met all three criteria. All four of the countries from Confucian Asia were represented by at least one sample.

In a number of studies, response bias was recognized by the researchers and bias type fit the pattern suggested in the discussion above - i.e., nomothetic items yielded extreme response bias on the part of the Asian sample and idiographic items led to midpoint or caution bias. It is important to note that investigators in these studies reported either a higher extreme response or higher midpoint bias for the Asian sample, but in none of these studies was any mention made of the idiographic-nomothetic distinction discussed above. As mentioned, the fact that the literature contains conflicting results has gone relatively unnoticed in this literature and the potential role of this factor thus far has not been recognized in the literature. Indeed, there has been relatively scant attention paid to any factors, aside from culture, which might play a role in the incidence of response bias.

For purposes of the current inquiry, the idiographic versus nomothetic determination for each study was made through descriptions of scales/items, or the items themselves, when given. It is worth noting that the failure to report actual items happens all too frequently in cross-cultural research. In a few studies surveyed here, all items were listed, in others, examples of items were offered, and in one case, it was possible to reasonably infer the nature of items.

Studies Employing Nomothetic items

Adler et al. (1989) set out in their study to examine attitudes of 103 Chinese managers and decided in the course of their study to abandon the original substantive focus entirely, based on the presence of a highly pronounced extreme response set. Items were nomothetic and involved statements about general principles and facts related to management practices requiring some level of agreement or disagreement. More than 55% of the Likert-type item response distributions were bimodal and at the scale extremes; very few middling opinions were offered. For 43 of 56 items, more than 60% of Chinese chose two categories at one end of the item scale.

Kotabe et al. (1991) used a sample of 689 American managers and 205 Japanese managers, to compare management perspectives of the two groups. Items consisted of 28 items making statements about strategic management principles (e.g., *"firms with high levels of brand awareness are more profitable"* and *"large companies benefit most from R&D efforts"*). Respondents were asked to respond on a five-point scale, ranging from *"almost never true"* to *"almost always true."* Items were thus nomothetic in content and format. The authors

mentioned the possibility that responses had been affected by either social acquiescence, or courtesy bias, thought to higher in parts of Asia, including Japan.

In their study, however, Kotabe et al. (1991) also cited the possibility of extreme response bias, which the authors held to be higher in American samples and noted that standard deviations in responses were statistically higher in half of the 28 items. The standard deviation evidence--despite its use in a number of studies--is less than convincing, however, for two reasons. First, it only applied to half of items, and second, because of the fact that it is entirely possible for standard deviations to be significantly greater for one group, without any extreme responses at all being chosen by that group. Interestingly, the comparisons of raw scores between the two groups, before standardization, showed that Japanese managers assigned greater assent than Americans did on 24 of the 28 items. When difference tests are used, this happens to be one of the prime effects of extreme response tendency--the group employing extreme response will be higher on all or almost all of the dimensions tested. It is difficult not to suspect that some combination of acquiescence and extreme response was operating on the part of Japanese managers, rather than on the part of American managers, as reported in the study--especially given the nomothetic nature of the items. This would comport with the folk knowledge principle held to be characteristic of respondents from countries that are historically Confucian.

Yates, et al. (1996) employed a sample of American and Taiwanese students in a study of decision-making overconfidence. Overconfidence was expected to be stronger for American students, but results showed greater overconfidence on the part of the Chinese. The authors explicitly related findings to the question of extreme response bias and cited findings as evidence for extreme response tendency on the part of the Chinese. They cited the phenomenon of Asian overconfidence, addressed in a number of studies in the decision-making literature, as the cultural rationale for the response set.

Finally, Culpepper et al. (2002) compared American and Chinese managers using items centered on reactions to various performance measures. Item ratings that were nomothetic engendered significantly more extreme responses on the part of Chinese respondents, while idiographic ratings and those involving specific informational content elicited a midpoint bias. Items included provided a fairly unique test of differing response tendencies, because the survey required three different types of ratings, nomothetic, idiographic and informational for each item stem. In other words item stems were held constant across the three types of ratings.

Studies Employing Idiographic items

In a study assessing across 53 countries the factor structure of the Rosenberg Self Esteem Scale, thoroughly idiographic, Schmitt & Allik (2005), the standard deviations for each of two subscales were reported for respective countries. As mentioned above, relative size of standard deviations has been used as a measure, despite its limitations, of extreme response in numerous studies, with a higher standard deviation being indicative of extreme response tendency. In this study, among standard deviations reported for Self-Competency and Self Liking subscales across 53 countries, each of the eight standard deviations from Confucian Asia - China, Hong Kong, Japan and South Korea -- were ranked at or near the bottom in size among the 53 countries. This suggests a robust tendency to gravitate to positions at or nearer the midpoint, also known as caution bias. We suggest that this may have stemmed from the use of idiographic item format.

Chun et al. (1974) employed a sample of 391 Korean and American students who answered surveys made up of idiographic items. Again, authors confirmed the presence of response bias and reported that Koreans used extreme responses less often than Americans. This is in line with the rationale discussed above, specifically, that respondents from countries that are part of Confucian Asia will revert to a midpoint or cautious bias when asked to agree or disagree with idiographic statements.

In another study employing idiographic items, Chen et al. (1995) using a sample of 5162 Taiwanese, Japanese, and Canadian students, reported that Taiwanese and Japanese students were less likely than two North American samples to use the midpoint. Parenthetically, they converted seven-point scales to three-point scales as a remedy.

Schermerhorn & Bond (1991), in a study comparing 101 Hong Kong and 102 American graduate students answering idiographic items, reported that a strong response bias prevented proper analysis using the original scores. Of the eight influence tactics that respondents were asked to agree/disagree with, Hong Kong means were lower on all eight comparisons, and a main effect for culture was discerned using MANOVA. The authors reported that in order to allow for more valid and conservative comparisons, raw scores of the Hong Kong respondents were increased by a constant of .12. This finding comports with a caution/midpoint bias for Hong Kong respondents faced with idiographic items.

Finally, a study by Zax & Takahashi (1967), in a comparison of Japanese and American subjects' cultural attitudes employed a novel type of item in which the primary referent was an inkblot. Inkblots were rated using a seven-point semantic differential scale, for example, ranging from "*wise to foolish*" to "*valuable to worthless*." This item format-insofar as there is no assertion or proposition in play clearly precludes any reversion to a sense of folk wisdom or "*what everybody knows*" that might stem from Confucian cultural influence. Consistent with such a rationale, Japanese subjects displayed a midpoint or "*caution*" bias, in line with our expectations for idiographic items.

Finally, a study comparing American and Taiwanese elementary school children Stigler et al. (1985) employed four attitudinal scales assessing perceived self-competence via idiographic items. Among 16 comparisons of means between American and Taiwanese school children, in only two comparisons was the Chinese mean response higher than the American mean response. Results were reported as completely substantive, although the researchers did mention the possibility that a self-effacing cultural norm may have been in play on the part of the Taiwanese children. In light of the above discussion, it is difficult to have confidence that results in fact stemmed completely from the substantive issues under study. One suspects, rather, that a midpoint/caution bias was in play given idiographic items and respondents from a Confucian culture.

SUMMARY

In the studies surveyed, the type of response tendency exhibited in Western-Asian comparisons did not hold across differing rating task types. The reviewed work offers valuable evidence to help reconcile divergent theory and findings in previous work, wherein some researchers have asserted a Confucian-oriented midpoint bias and others have claimed an extreme response bias, stemming from the Confucian history, for the very same countries. Much more work is needed to further clarify and resolve this ambiguity.

The observed patterns in previous work do not offer definitive conclusions about the cultural factors cited here as the source of these response bias differences. On the other hand, existing theory about cultural influences is helpful in guiding expectations about the role that item type may play in the nature of response bias exhibited by different cultural groupings.

It is important to note that empirical patterns highlighted in this survey are important in and of themselves, regardless of what theory may explain such patterns. This is because some have argued that response bias is a mere artifact of test-taking style, having no substantive significance (Chun et al., 1974). Our analysis renders such an explanation much less tenable, because a mere style, whether toward the extremes or midpoint, would have applied more or less equally across idiographic and nomothetic types of multiple-response rating tasks. This was not the case, suggesting that substantive cultural factors are in play—regardless of what their precise nature may be. Furthermore, establishing the link between item type and bias type for Confucian countries should be helpful in guiding future research as to which response tendencies to expect and to be ready to handle methodologically, as well as, in some cases, guiding choice of items. Many of the studies cited here stumbled upon the response bias issue on an incidental basis. Perhaps more work in this area, will better prepare future cross-cultural work in attuning researchers to this problem from the study's outset.

REFERENCES

- Adler, N.J., Campbell, N., & Laurent, A. (1989). In search of appropriate methodology: From outside the People's Republic of China looking in. *Journal of International Business Studies*, 20(1), 61-74.
- Chen, C., Lee, S.Y., & Stevenson, H.W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6(3), 170-175.
- Chun, K.T., Campbell, J.B., & Yoo, J.H. (1974). Extreme response style in cross-cultural research: A reminder. *Journal of Cross-Cultural Psychology*, 5(4), 465-480.
- Cronbach, L.J. (1946). Response sets and test validity. *Educational and psychological measurement*, 6(4), 475-494.
- Cronbach, L.J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10, 3-31.
- Culpepper, R.A., Zhao, L. & Lowery, C. (2002). Survey response bias among Chinese managers. *Academy of Management Proceedings*, 2002(1), J1-J6.
- Hamilton, D.L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, 3, 192-203.
- Hui, C.H., & Triandis, H.C. (1989). Effects of cultural and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20, 296-309.
- Jaccard, J., & Wan, C.K. (1986). Cross-cultural methods for the study of decision-making. *Journal of Cross-Cultural Psychology*, 17, 123-149.
- Kotabe, M., Duhan, D.F., Smith, D.K., & Wilson, R.D. (1991). The perceived veracity of PIMS strategy principles in Japan. *Journal of Marketing*, 55, 26-41.
- Lee, C., & Green, T. (1991). Cross-cultural examination of the Fishbein behavioral intentions model. *Journal of International Business Studies*, 22, 289-305.
- Lee, J.W., Yates, J.F., Shinotsuka, H., Singh, R., Ongiatco, M.L.U., Yen, N.S., Gupta, M., & Bhatnagar, N. (1995). Cross-national differences in overconfidence. *Journal of Asian Psychology*, 1, 63-69.
- Leung, K., & Bond, M.H. (1989). On the empirical identification of dimensions for cross-cultural comparisons. *Journal of Cross-Cultural Psychology*, 20, 133-151.
- Mullen, M. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Research*, 573-597.
- Schermerhorn, J.R., & Bond, M.H. (1991). Upward and downward influence tactics in managerial networks: A comparative study of Hong Kong Chinese and Americans. *Asia Pacific Journal of Management*, 8(2), 147-158.

- Schmitt, D.P., & Allik, J. (2005). Simultaneous Administration of the Rosenberg Self-Esteem Scale in 53 Nations: Exploring the Universal and Culture-Specific Features of Global Self-Esteem. *Journal of Personality and Social Psychology*, 89(4), 623–642.
- Stening, B.W., & Everett, J.E. (1984). Response styles in a cross-cultural managerial study. *Journal of Social Psychology*, 122, 151-156.
- Stigler, J.W., Smith, S., & Mao, L. (1985). The Self-Perception of Competence by Chinese Children. *Child Development*, 56(5).
- Whitcomb, K.M., Oenkal, D., Curly, S.P., & Benson, P.G. (1995). Probability judgment accuracy for general knowledge: Cross-national differences in assessment methods. *Journal of Behavioral Decision-Making*, 8, 51-67.
- Wright, G.N., & Phillips, L.D. (1980). Cultural variation in probabilistic thinking: Alternative ways of dealing with uncertainty. *International Journal of Psychology*, 15, 239-257.
- Yates, J.F., Lee, J.W., & Bush, J.G. (1997). General knowledge overconfidence: cross-national variations, response style, and “reality”. *Organizational Behavior and Human Decision Processes*, 70(2), 87-94.
- Yates, J.F., Lee, J.W., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes*, 65(2), 138-147.
- Yates, J.F., & Lee, J.W. (1996). Chinese decision making. *Hong Kong: Oxford University Press*.
- Yates, J.F., Ji, L.J., Oka, T., Lee, J.W., Shinotsuka, H., & Sieck, W.R. (2010). Indecisiveness and culture: Incidence, values, and thoroughness. *Journal of Cross-Cultural Psychology*, 41(3), 428-444.
- Yates, J.F., Zhu, Y., Ronis, D.L., Wang, D.F., Shinotsuka, H., & Toda, M. (1989). Probability judgement accuracy: China, Japan, and the United States. *Organizational Behavior and Human Decision Processes*, 43, 147-171.
- Yu, J.H., Keown, C.F., & Jacobs, L.W. (1993). Attitude Scale Methodology: Cross-cultural implications. *Journal of International Consumer Marketing*, 6, 45-64.
- Zax, M., & Takahashi, S. (1967). Cultural influences on response style: Comparisons of Japanese and American college students. *Journal of Social Psychology*, 71, 3-10.
- Zhao, L. & Culpepper, R. (1996). Performance measures for China-based joint ventures: A comparison of Chinese and American managers. *Journal of Asian Business*, 12, 33-50.

Received: 02-Nov-2022, Manuscript No. JIBR-22-12992; **Editor assigned:** 04-Nov-2022, Pre QC No. JIBR-22-12992(PQ); **Reviewed:** 18-Nov-2022, QC No. JIBR-22-12992; **Revised:** 25-Nov-2022, Manuscript No. JIBR-22-12992(R); **Published:** 30-Nov-2022