

1-1-2013

The equivalence of Internet versus paper-based surveys in IT/IS adoption research in collectivistic cultures: the impact of satisficing

Jiaming Fang

University of Electronic Science and Technology of China

Chao Wen

Eastern Illinois University, cwen@eiu.edu

Victor Prybutok

University of North Texas

Follow this and additional works at: http://thekeep.eiu.edu/business_fac



Part of the [Business Commons](#)

Recommended Citation

Fang, Jiaming; Wen, Chao; and Prybutok, Victor, "The equivalence of Internet versus paper-based surveys in IT/IS adoption research in collectivistic cultures: the impact of satisficing" (2013). *Faculty Research & Creative Activity*. Paper 5.
http://thekeep.eiu.edu/business_fac/5

This Article is brought to you for free and open access by the Business, School of at The Keep. It has been accepted for inclusion in Faculty Research & Creative Activity by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.

The equivalence of Internet versus paper-based surveys in IT/IS adoption research in collectivistic cultures: the impact of satisficing

Jiaming Fang, Chao Wen, & Victor Prybutok

Behaviour & Information Technology

An increasing proportion of information technology (IT)/information system adoption research collects data using online surveys. However, a paucity of research assesses the equivalence of paper-based versus Internet-based surveys in collectivistic cultures. Furthermore, no theoretical or empirical research investigates how cultural differences between collectivistic and individualistic cultures influence the measurement equivalence (ME) of these survey modes. To explore these issues, online and paper-based surveys with comparable samples were carried out in both an individualistic (the USA) and a collectivistic culture (China). Confirmatory factor analysis was conducted to examine the ME across both survey modes in these different cultures. Results indicate that the relatively larger satisficing discrepancy between paper and online surveys causes respondents in collectivistic cultures to have an increased likelihood of providing responses that vary as compared to respondents in individualistic cultures. The disparate responses, in turn, result in increased measurement variance between the two survey modes. The findings of this study bridge a gap in the literature and address the question of how culture influences online satisficing behaviour and how that behaviour causes measurement invariance across survey modes. This study also explains the possible underlying mechanisms by which different national cultures exert their influence on survey results. The findings provide important implications for IT researchers, especially those in collectivistic cultures or those who need to collect data in collectivistic cultures using online surveys or mixed-mode surveys that include an online survey mode.

1. Introduction

The increasing use of information technology (IT) has amplified the importance of research on IT/information system (IS) adoption and decision-making behaviour (e.g. Behrend et al. 2011, Green and Pearson 2011, Zhou 2011). Abareshi and Martin (2009) conducted a meta-analysis on three IS journals (MISQ, JMIS, and I&M) over the past 15 years to investigate the methods used in these fields. They found that survey research is prominent and an increasing proportion of that research was conducted using Internet surveys (e.g. Kim et al. 2009, Setterstrom et al. 2012).

The use of the Internet to collect data is constrained by coverage limitations and the participants' lack of willingness to respond. Numerous studies were conducted to identify the reasons that participants do not respond to Internet surveys and to develop approaches for increasing the willingness of potential respondents to

participate in Internet surveys (e.g. Bosnjak et al. 2005, Heerwegh and Loosveldt 2007, Fang et al. 2012a). Despite the inaccuracies caused by low coverage, few studies have addressed the coverage limitation issue. One solution proposed to alleviate the negative effect of low coverage is to use a mixed-mode survey design for data collection. For example, researchers can use an Internet survey to collect data and then access additional respondents by collecting data via another delivery mode such as a paper-based survey. This approach provides sufficient data for analysis since the advantages of one survey mode can partially compensate for the disadvantages of the other survey mode.

However, when integrating data collected from Internet surveys with traditional paper-based surveys, researchers and practitioners must ensure the measurement equivalence (ME)¹ of these two modes. ME is defined as the extent to which measurements conducted under different conditions measure the same attributes (Miles and King 1998). Mixedmode surveys can serve as a trustworthy basis for decisionmaking provided that convergent validity between the two survey methods is established (Deutskens et al. 2006).

Numerous measurement invariance (MI) studies were conducted to examine the equivalence of Internet surveys with their paper counterparts (e.g. King and Miles 1995, Booth-Kewley et al. 2007, Meade et al. 2007), but most of this research was conducted within individualistic cultures. Few empirical studies have investigated whether paper and online surveys produced convergent results in collectivistic cultures. One exception is the work of Fang et al. (2012b) that examined MI in collectivistic cultures but that work did not compare the difference in MI in collectivistic and individualistic cultures. As a result, it is not clear if culture has an influence on MI. Furthermore, no theoretical or empirical research has so far explored the process or underlying mechanisms responsible for the effect of culture on ME. Further, most existing MI studies are conducted in research areas such as psychology or public opinion. These studies mainly focus on variables such as personality, depression, and IQ. However, in the research domains of IT/IS adoption, a substantial body of the research involves measurements about attitudes and self-assessments. These IT-related measurements involve variables which are more likely to be culture dependent than those in the research domains of psychology or public opinion. Consequently, investigating MI issues in the IT/IS adoption research domain in collectivistic cultures has the potential to provide valuable insights for researchers. This is an important research topic precisely because little effort has been devoted to the area.

To address this gap in the IT literature, comparable online and paper-based surveys were conducted in both the USA (with an individualistic culture) and China (with a collectivistic culture). Multi-group confirmatory factor analysis (MCFA) was performed to evaluate the ME of the personal innovativeness scales, which are among the most frequently used scales in the IT/IS adoption domain. We expected that the lean nature of the Internet and the dissimilar Internet consumption patterns between individualistic and collectivistic cultures would impact the ME. Specifically, we posit that the cultural differences would make the ME across survey modes more likely to be violated in collectivistic societies. Further, we explored the potential mechanisms responsible for the measurement variance.

This study contributes to the wide range of IT/IS adoption research and survey research by offering empirical evidence for MI between Internet and paper-based surveys in collectivistic cultures. More importantly, the present study endeavours to identify the potential mechanism underlying the effect of culture on measurement variance. We believe that this expanded focus will broaden the range for which these results are applicable beyond those currently investigated and therefore provide valuable implications for IT/IS researchers, especially those collecting data in collectivistic cultures.

The rest of this paper is organised as follows. First, we review the literature on MI between online and paper-based surveys, the satisficing impacts on MI and the moderation effect of cultural differences on satisficing. Based on this literature review, we propose our research questions. After that, we present the research methodology and results of the data analysis. Finally, we discuss the theoretical and practical implications of our findings, the limitations, and some recommendations for future research.

2. Background

2.1. MI between online and paper-based survey modes

To integrate the data collected from Internet surveys with data from paper-based surveys, one must assume that the data gathered using these two survey modes are comparable. In other words, the surveys should have MI across these two different survey modes and that the numerical values under consideration are measured on the same scale. A failure to demonstrate ME across survey modes not only casts doubt on their comparability but also makes unambiguous interpretation impossible (Vandenberg and Lance 2000, Cole et al. 2006).

There are several existing studies regarding the comparison of online and paper-and-pencil survey modes. Some of these studies found that there is no evidence that the administration mode affects ME (e.g. King and Miles 1995, Leidner and Kayworth 2006), while other studies indicate that survey modes indeed can affect the results (e.g. Fouladi et al. 2002, Buchanan et al. 2005, Meade et al. 2007). Although the results remain inconclusive, these studies have contributed to our understanding of MI between paper and online surveys.

Although a number of empirical studies were conducted to examine the differences in observed scores in online surveys and paper-based surveys, most of these studies (e.g. Braunsberger et al. 2007, Stephenson and Crête 2011) only compared mean scores across administration modes and did not examine the differences using an MI framework. Because means are just approximations of the underlying true population means, it is impossible to claim how much importance should be attached to small but significant differences (Deutskens et al. 2006).

In addition, almost all the MI studies examining the equivalence of Internet-based and paper-based survey modes were carried out in individualistic cultures. It is not clear whether there are cultural impacts on MI because of the lack of cross cultural comparisons. Furthermore, it is not a given that the equivalence of different

survey methodologies in individualistic cultures such as the USA will remain true when transferred to other cultures.

2.2. The impact of satisficing on MI

IT/IS researchers very often resort to self-administered attitude scales such as the perceived usefulness, perceived ease of use, and perceived enjoyment in studying adoption behaviours. Krosnick (1991) proposes a theory called satisficing to interpret a variety of respondent behaviours involving attitude measurements. According to the contemporary model of answering survey questions, survey respondents must implement four stages of cognitive processing to answer survey questions optimally: interpret the intended meaning of the question, retrieve relevant information from memory, use this information to make required judgements, and select and report an answer (Kaminska et al. 2010, Tourangeau et al. 2000). When respondents expend a great deal of cognitive effort in each of these four steps, they are said to be optimising. But if the cognitive demands of doing this work exceed the respondent's motivation or ability, a respondent may choose to perform each of the steps in a less diligent and comprehensible way (weak satisficing), or they may simplify their endeavour by skipping one or more steps altogether, and just superficially interpret each survey question instead of thinking carefully about their response (strong satisficing). Satisficing is especially present in attitude measurement or answering a survey question when substantial cognitive effort is required. A recent study shows that the majority of respondents in the USA engage in satisficing behaviours in surveys which can distort parameter estimations by increasing scale reliabilities and altering statistical relationships between constructs (Barge and Gehlbach 2012). For instance, satisficing behaviours can enable the otherwise uncorrelated relationship between two scales to be significantly correlated.

In mixed-mode survey research, which consists of online surveys and paper surveys, we speculate that respondents to online surveys are less likely to expend the required cognitive effort to answer questions carefully compared with the respondents to paper surveys. As a result, the respondents to online surveys should exhibit more satisficing, thereby leading to measurement nonequivalence between the two survey modes. There are two major rationales underlying the hypothesis. First, compared to online surveys, social distance between respondents and the survey researcher is relatively less in offline paper surveys. As a lean medium, the Internet provides few social ties, especially in ad hoc Internet surveys, where follow-up surveys are not conducted on the same sample. Therefore, the relationships between participants and survey researchers are temporary and weak. The presence of a survey researcher in a paper survey may help minimise the extent to which respondents are distracted, and the survey researcher's presence offers the possibility of developing interpersonal trust (Roberts 2007). Second, online surveys offer respondents a greater opportunity for multitasking than in paper surveys (de Leeuw 2005). Online survey respondents are more easily distracted by such things as listening to music, watching videos, chatting with friends, and playing online games when taking the survey. Thus, the respondents

may not devote adequate cognitive effort to answering the survey questions. Compared to an offline environment, the Internet easily creates a social situation that produces a sense of disinhibition, and this sense of disinhibition increases the likelihood that respondents will engage in multitasking.

2.3. The impact of national culture on satisficing

We posit that the respondents in a collectivistic culture are more likely to engage in a greater degree of satisficing in online surveys than when the survey researcher is present versus their counterparts in an individualistic culture. Culture is the collective programming of the mind distinguishing the members of one group or category of people from others (Hofstede et al. 2010). Culture shapes the definition of acceptable behaviours through the norms and values of society. Instead of examining the whole range of cultural dimensions, this study focuses on individualism–collectivism. Individualism–collectivism is believed to be the most important construct for explaining the social psychological patterns of cultures (Levine et al. 2001), and it is also the most commonly used dimension to understand the differences between cultures (Calhoun et al. 2002, Gaspay et al. 2008). In an individualistic culture (e.g. the USA, Australia and most of Western Europe), individuals are more concerned with the self, social ties between individuals tend to be loose, and communication is relatively direct (Triandis 1988). In individualistic cultures, an individual is an independent entity and possesses relatively more free will, emotions, and personality. People in individualistic cultures tend to make decisions and to initiate behaviours independent of others (Bao et al. 2003). Alternatively, in a collectivistic culture (e.g. China, Malaysia, and Singapore), individuals stress their social needs more than their private needs when making decisions.

The following three features contribute to the likelihood of engaging in more salient satisficing behaviour on the Web in a collectivistic culture. First, face consciousness enables collectivists to be more sensitive to the evaluations of others compared to their counterparts in individualistic cultures. Face consciousness refers to people's desire to enhance, maintain, and avoid losing face in relation to other people in social activities (Bao et al. 2003). Compared to individualists, people in collectivistic cultures are more concerned about face consciousness and it plays an important role in collectivistic countries (Keh and Sun 2008). Collectivistic societies celebrate the values of fulfilling one's obligations and responsibilities over one's own personal wishes or desires. In many paper survey formats, respondents will devote sufficient mental effort to the questionnaire because the survey researchers and respondents have physical contact. Failing to appropriately engage and participate in a survey is not an acceptable social value in a collectivistic culture. In contrast, because of the impersonal online environment and the social anonymity, individuals perceive much less social pressures in online contexts, but respondents will still not refuse to participate in surveys since in collectivistic cultures a direct confrontation is always avoided. Tan et al.'s (1998) study supported this claim because they found that collectivists could not resist the influences of their culture even on the Web.

The second reason is that to participate in online surveys is incongruous with respondents' Internet consumption motivation. Both respondents in collectivistic cultures and individualistic cultures may similarly engage in multitasking on the Web. However, the primary motivation for Internet usage in collectivistic societies is hedonic and social-oriented, while the Internet usage motivation is more utilitarian in individualistic societies. La Ferle and Kim (2006) examined cultural influences on Internet motivations and the results showed that US citizens were more likely to be motivated to shop online and look for information online than Korean consumers (collectivists). In contrast, Korean consumers were more motivated than US consumers to head online to fulfil social goals such as making friends, meeting people and participating in newsgroups. Ko et al. (2006) further confirmed this conclusion in their study. Ko et al. (2006) found that individualists had a higher degree of information and convenience motivation and perceived a higher degree of human-message interaction, while collectivists had a higher degree of social interaction motivation and human-human interaction. Participation in online surveys is inconsistent with the online entertainment motivations of respondents in collectivistic cultures. The respondents in collective cultures find it difficult to focus on taking part in surveys when engaging in other social activities such as chatting with friends, playing online games, and watching online movies. However, very often they will not reject the survey participation request directly, because such rejection disrupts interpersonal harmony and is not encouraged by their cultural values. As a result of these two issues, respondents will devote insufficient effort to completing the online surveys and the possibility of satisficing is higher than for the respondents in individualistic cultures.

Finally, the difficulty of developing online trust in collectivistic cultures leads to more salient satisficing behaviours. Previous research shows that individualists are less concerned with self-categorising, are less influenced by group membership, and are more ready to trust others than collectivists in online environments (Jarvenpaa and Leidner 1999). In an online context, where social ties are minimal, building trust in a collectivist culture is more difficult than in an individualist culture. In collectivistic cultures there is relatively greater concern for the collective and less attention is given to outsiders than in individualistic cultures (Levine et al. 2001). Thus, it is speculated that collectivists are less likely to expend a great deal of effort filling in online questionnaires than their counterparts in individualistic cultures.

Given the impact of collectivism-individualism on satisficing on the Web, we infer that compared with individualistic cultures, MI across survey modes is more likely to be violated in collectivistic cultures.

3. Methodology

3.1. Operationalisation of satisficing and national culture

Hofstede et al. (2010) provides individualism-collectivism scores for 76 countries and regions. The USA and China have scores of 91 (individualistic) and 20 (collectivistic), respectively. Thus, the differential impact on survey modes between

individualistic and collectivistic cultures was assessed using matching online and paper-based surveys in both the USA and China.

There is no existing scale to measure the level of satisficing. Previous research has normally measured satisficing using a single indicator such as middle response (Krosnick et al. 1996), extreme response (Belli et al. 1999) and numbers of 'don't know' responses (Kaminska et al. 2010). However, satisficing is a multifaceted construct (Kaminska et al. 2010). In this study, we assessed satisficing using two metrics. An 'extreme + middle' metric was used to count the number of times a respondent chose 1, 4, or 7 in response to questions with a 7-point Likert scale. A second metric of inconsistent responses was used to gauge the extent of answer variation within a respondent.

3.2. Scale selection

The objectives of the present study make it necessary to exclude confounding factors such as scales, which can also potentially influence the results of the MI assessment. Furthermore, prior research suggested that variation in satisficing effects by mode might also be dependent on additional factors such as the length of the questionnaire and the types of questions asked (Holbrook et al. 2003, Fricker et al. 2005, Roberts 2007). Thus, we used the same scales for both the online and paper surveys. Finally, we selected two scales to assess MI issues. One is the personal innovativeness in the domain of IT (PIIT) as proposed by Agarwal and Prasad (1998). PIIT is used to describe willingness to adopt a new IT (Agarwal and Karahanna 2000, Lu et al. 2005) and is a frequently used scale in IT/IS adoption research. The PIIT scale was used worldwide with the evidence consistently showing that PIIT is an important predictor of individual IT/IS adoption behaviour. The second scale is Global Innovativeness (GI), which assesses an individual's general predisposition to innovativeness, or willingness to try new things, as proposed by Hurt et al. (1977). This scale is frequently adopted in marketing research (Clark and Goldsmith 2006) and has IT/IS applications. Given that PIIT belongs to the domain of a specific innovativeness scale, we posit that a significant positive correlation exists between PIIT and GI. The original scales are in English, and they were translated into Chinese with translation/back translation techniques.

3.3. Data collection

Undergraduate students from a large southwest public university in the USA were chosen for data collection. To ensure the quality of the experiment, we selected two sections of students taking the same introductory level business statistics course as a sample for the US data. Each section has a class capacity of 200 students. From the paper survey group, we received 156 responses. The final sample included 150 completed and valid questionnaires. The sample consisted of 97 men and 50 women, and 3 respondents did not provide their gender information. The majority of the respondents (93.9%) were between 18 and 28 years of age. In the online survey group, 149 completed questionnaires were obtained. After eliminating invalid responses, 140 useful respondents remained in

the sample. The sample consisted of 77 males and 63 females. Respondents in both groups had a similar Web experience, gender composition, and educational level.

In China, we adopted a similar method for collecting data. Students from a large public university in the southwest of China were recruited to participate in both the paper-based and online surveys. We distributed 200 paperbased questionnaires in classes, and received 180 responses. The final sample included 139 complete and valid questionnaires. The sample consisted of 94 men and 45 women. The majority of the respondents (95.7%) were between 18 and 28 years of age. An independent student sample was chosen for the online survey. The Web-based survey questionnaire was hosted on a well-established online survey hosting site (<http://www.sojump.com>) in China. The site is similar to SurveyMonkey.com or Qualtrics.com in the USA. E-mail addresses of the potential respondents were obtained in advance from classes and they were informed by the researchers that an email invitation in which a survey URL was embedded would be sent to them. A total of 228 emails were sent, and in 2 weeks we received 196 responses of which 137 responses were valid. The sample consisted of 59 males and 78 females. Most of the respondents (81.2%) were between 18 and 28 years of age. Respondents in both groups had a similar Web experience and educational levels, but the gender composition was different.

To evaluate the demographic similarity between the USA and China samples, we conducted a series of chi-square tests on age, gender composition, and Web experience. Results indicated that there were no significant differences in these variables between the USA and China samples at the 5% level.

To verify the adequacy of the sample size to conduct multi-group confirmatory analysis (MCFA), this investigation adopted Westland's (2010) methodology and calculated the ratio of indicators to latent variables. The results showed that the samples were adequate for conducting the confirmatory factor analysis (CFA).

3.4. Testing procedures

We employed MCFA. Consistent with previous research (Vandenberg and Lance 2000, Vandenberg 2002), we first examined the equality of the observed variance-covariance matrices. If the matrices were invariant across the survey modes, additional analyses to assess MI were unnecessary. In this case it is appropriate to pool the data from the different survey modes. Otherwise, we assessed the equality of a number of factors and factor pattern matrices (configural invariance or weak factorial invariance) to assess if the assumption of equality in the covariance matrices was not satisfied (Raju et al . 2002). The configural invariance model was viewed as a baseline model against which further invariance models were evaluated. Next, we tested the invariance of the factor loadings matrix, and then we evaluated the invariance of item intercepts and the covariance matrix (phi matrix) among the factors across the survey modes.

The chi-square goodness of fit and three index measures (the Tucker-Lewis index (TLI), comparative fit index (CFI), and root mean square error of approximation (RMSEA)) provided the general model fit measurements to

assess goodness of fit. These indices are sensitive to a lack of MI and were frequently used in past MI research (e.g. Vandenberg and Lance 2000, Leidner and Kayworth 2006, Chen 2007). Although the chi-square difference statistic was the primary criterion, the statistic is sensitive to sample size and violation of the multivariate normality assumption (Chen 2007). As rules of thumb the TLI and CFI values should be greater than 0.90 to indicate a good fit and the RMSEA value should be less than 0.08 (Hair et al . 1998).

4. Data analysis and results

4.1. Descriptive data analysis

We evaluated internal consistency using Cronbach's alpha. In the group of paper-based surveys, the GI and PIIT alphas were 0.85 and 0.75, respectively. The corresponding alpha coefficients were 0.82 and 0.75 in the website-based survey group. These reliability coefficients were all above the suggested cut-off value of 0.7 (Nunnally and Bernstein 1994).

We evaluated the reliability of our constructs in each sample using composite reliability and average variance extracted (AVE). The composite reliability scale values ranged from 0.86 to 0.90 with all values exceeding the cutoff point of 0.70. The AVE ranged from 0.54 to 0.75, so all constructs exceeded the 0.50 cut-off value. Further, we assessed discriminant validity and we compared the AVE of each construct to the square of the correlations among the latent variables to evaluate the discriminant validity. Results showed that the AVE of each construct was larger than the square of the correlations among the latent variables, which supported the discriminant validity for each construct.

The Doornik–Hansen tests for multivariate normality showed that there were significant deviations from multivariate normality both in the US sample ($\chi^2(14) = 218.94, p < 0.01$) and in the China sample ($\chi^2(14) = 127.89, p < 0.01$). Thus, we used a robust maximum likelihood (ML) estimator with the Yuan–Bentler corrections for non-normal data instead of the standard ML estimator.

4.2. MI assessment

4.2.1. Test of the variance–covariance matrices invariance

Initially, we assessed the equality of the covariance matrices and mean vectors of the observed items in the US sample. Our results showed that there were no significant differences in the covariance matrices between online and paper survey modes in the US sample, $\chi^2(28) = 36.57, p = 0.13$. Testing for the differences in the means also supported the model fit, $\chi^2(7) = 8.61, p = 0.28$. Pooling the data from the online and paper surveys is appropriate because the covariances and means were invariant across the survey modes and additional MI tests are redundant (Vandenberg and Lance 2000, Vandenberg 2002, Deutschens et al. 2006).

The covariance matrices and means were examined using the China sample. The results showed that $\chi^2(28) = 43.05$, $p = 0.03$. However, testing for the equity of the mean vector allowing heterogeneous covariance matrices across groups resulted in insignificant results, $F(7,265.9) = 1.95$, $p = 0.06$. In sum, the results rejected the null hypothesis that variance–covariance matrices were invariant across the two survey modes, so it was necessary to identify the source of nonequivalence (Vandenberg and Lance 2000). Subsequent MI tests were conducted only on the China sample.

4.2.2. Test of configural invariance

To determine whether the equality of the number of factors and factor pattern matrices hold in each survey mode, we conducted an exploratory factor analysis. The results indicated that seven items loaded onto two factors (GI and PIIT) and cross-loadings were well below the usual cutoff value of 0.40. This model (M1) provided the baseline model value against which subsequent tests for invariance were compared. The results are listed in Table 1.

Table 1. Test of configural invariance.

Model	χ^2_{YB}	df	TLI	CFI	RMSEA
M1	45.35	26	0.93	0.96	0.07

Note: χ^2_{YB} , the Yuan–Bentler scaled chi-square statistic.

The fit indices show that we could not reject the hypothesis that the equality of the number of factors and factor pattern matrices were consistent in each survey mode.

4.2.3. Test of the invariance of the factor loadings matrix

However, because configural invariance was established, we further evaluated MI by constraining the factor loadings to invariance across survey modes (M2). The results are shown in Table 2. We adopted the method proposed by Satorra and Bentler (2001) to calculate the chi-square difference.

In Table 2, after constraining the factor loadings to invariance across survey modes, the nested model still fits well and the change in chi-square was not significant ($p = 0.09$). Therefore, metric invariance was established.

Table 2. Results of the invariance of factor loadings matrix assessment.

Model	χ^2_{YB}	df	TLI	CFI	RMSEA	$\Delta\chi^2$	Δdf	<i>p</i>
M2	54.80	31	0.93	0.95	0.08	9.39	5	0.09

Note: M2 constrains the factor loadings to invariance across the survey modes.

4.2.4. Test of the invariance of the item intercepts across survey modes

To further assess whether the two survey modes possessed the same item intercepts or scalar invariance, we constrained the item's intercept equality across survey modes based on M2 (M3). When intercept invariance is achieved, it indicates that scores from different groups have the same origin because the intercept is the origin of the scale (Chen 2007). M3 was nested within M2. As shown in Table 3, an insignificant chi-square change indicated that the intercept equality assumption was supported.

Table 3. Results of the invariance of item intercept assessment.

Model	χ^2	df	TLI	CFI	RMSEA	$\Delta\chi^2$	Δdf	<i>p</i>
M3	66.32	38	0.93	0.94	0.07	11.41	7	0.12

Note: Model M3 constrains loadings and item intercepts to invariance across survey modes.

4.2.5. Test of the invariance of the covariance matrix among latent variables

To assess the invariance of the covariance matrix we constrained the covariance to equality based on Model 3 while keeping the variances of the factors freely estimated (M4.1). Because M4.1 was nested within M3 we could assess the model fit compared to the M3. The results are shown in Table 4. The comparison yielded a chi-square difference value of 10.98 with 1 degree of freedom, which was statistically significant ($p < 0.01$). Data analysis revealed that the covariance of the latent variables in the paper survey mode (0.23) was smaller than that in the online survey mode (0.61), even though the covariances in both groups were significant.

Because Model M4.1 was rejected, we respecified the model by fixing the variances of the respective factors to invariance while relaxing the covariance (M4.2). M4.2 was also nested in M3. The results in Table 5 show an insignificant chi-square change and support the variances of the factors equality assumption.

Table 4. Results of the invariance of covariance matrix assessment.

Model	χ^2_{YB}	df	TLI	CFI	RMSEA	$\Delta\chi^2$	Δdf	<i>p</i>
M4.1	74.06	39	0.92	0.93	0.08	10.98	1	< 0.01

Note: M4.1 constrains the loadings, intercept and covariances to invariance across survey modes.

Table 5. Results of the invariance of variance matrix assessment.

Model	χ^2_{YB}	df	TLI	CFI	RMSEA	$\Delta\chi^2$	Δdf	<i>p</i>
M4.2	65.77	40	0.94	0.95	0.07	0.95	2	0.62

Note: M4.2 constrains loadings, intercepts, and variances to invariance across survey modes.

4.3. Testing the impact of satisficing on MI

The MI tests showed that the measurement nonequivalence in the China sample was the result of the covariance differences. Since the variance of the latent variables was equivalent across survey modes, the larger covariance in the online survey implied a closer correlation between the latent variables, which was probably caused by similar scorings on the indicators. Previous research shows that satisficers do not select the exact same response for items but insufficiently adjusted their responses from one item to the next and such response patterns result in inflated correlation between the two scales (Barge and Gehlbach 2012). We believed that the disparate associations also originated from the satisficing behaviour of respondents. Thus, we further examined whether the levels of satisficing differed in paper and online surveys and whether there was significant disparity between satisficing in the USA and China samples. Considering that gender might be associated with satisficing, we included gender as one of the independent variables.

We first assessed whether there was an ‘extreme + middle’ scoring difference in offline surveys or online surveys between the USA and China samples. The results showed that in both paper and online surveys the interaction of culture and gender was insignificant (paper survey: $F(1,282) = 0.84$, $p = 0.36$; online survey: $F(1,273) = 0.05$, $p = 0.82$). The main effects of culture (paper survey: $F(1,282) = 0.01$, $p = 0.91$; online survey: $F(1,273) = 0.20$, $p = 0.65$) and gender (paper survey: $F(1,282) = 0.41$,

$p = 0.24$; online survey: $F(1,273) = 0.76$, $p = 0.38$) were also not significant.

Next, we examined whether there was a significant 'extreme + middle' scoring difference in offline and online surveys within a culture. We found that the interaction effects of survey mode by gender exerted an insignificant effect (the US sample: $F(1,283) = 0.32$, $p = 0.57$; China sample: $F(1,272) = 0.29$, $p = 0.59$) on 'extreme + middle' scorings. The main effect of gender was also not significant in both samples (the US sample: $F(1,283) = 1.40$, $p = 0.24$; China sample: $F(1,272) = 0.67$, $p = 0.41$). However, a significant main effect of survey mode was found (the US sample: $F(1,283) = 5.59$, $p = 0.02$, Cohen's $d = 0.26$; China sample: $F(1,272) = 13.31$, $p < 0.01$, Cohen's $d = 0.44$). The survey mode results show that respondents completing the online survey manifested significantly higher satisficing than those completing the paper survey in both China and the USA. Furthermore, the effect size of the survey mode in the US sample was a small effect, while the effect size in the China sample was a medium-to-large effect (Cohen 1988). These results show that respondents in China demonstrated a relatively larger satisficing discrepancy between the online and offline survey modes compared to their counterparts in the USA.

We computed the variance of response items within a respondent while excluding the items in which the respondent picked extreme values of 1 or 7 in response to questions 2 and used it to gauge the extent of variation in the responses. Our results showed that there was no significant interaction effect for gender by survey mode in either sample (the US sample: $F(1,261) = 3.14$, $p = 0.08$; China sample: $F(1,267) = 0.23$, $p = 0.63$). In the US sample, the main effect of gender was insignificant ($F(1,261) = 1.28$, $p = 0.26$), but the effect of survey mode was significant ($F(1,261) = 6.11$, $p = 0.01$, Cohen's $d = 0.25$), with online survey respondents offering relatively smaller variation in their answers. However, the main effects of gender and survey mode were both significant in the China sample. Responses provided by females in China showed significantly more variation than those of males ($F(1,267) = 6.23$, $p = 0.01$, Cohen's $d = 0.18$) and the variation among responses in the online survey was much smaller than in the offline survey ($F(1,267) = 15.09$, $p < 0.01$, Cohen's $d = 0.40$).

When we compared the inconsistent responses between the US samples and the China samples, the results showed that for both the paper and online surveys the US respondents provided more consistent answers. Furthermore, this consistency difference was statistically significant (paper survey: $F(1,272) = 30.25$, $p < 0.01$, Cohen's $d = 0.69$; online survey: $F(1,256) = 22.00$, $p < 0.01$, Cohen's $d = 0.61$).

Integrating the above results about 'extreme + middle' scorings and inconsistent responses, it is evident that respondents in the US sample were more likely to provide consistent responses between paper surveys and online surveys than respondents in the China sample. This difference in consistency held even if respondents in the US sample were potentially more liable to engage in satisficing behaviour on both surveys. It is the relatively large differences in responses between paper and online surveys that resulted in measurement variance in the China samples.

5. Discussion

In the IT/IS adoption and decision-making behavior research, both academics and practitioners have gradually switched to mixed-mode surveys. This shift to mixed modes is gradually replacing the traditional unimode survey methodology because it allows the researcher to obtain a larger number of representative responses with lower survey costs. However, few researchers have assessed MI when mixed-mode survey methods were used to collect data or considered the potential influences that mixed survey modes might have on the conclusions. In addition, almost all MI studies to date were carried out in individualistic cultures and little empirical evidence exists on MI in a collectivistic culture. Previous research indicates that a collectivistic culture and Confucian values can influence behaviour patterns (Leidner and Kayworth 2006, Zhang et al . 2009). This is an important area of research because it is possible that the equivalence of different survey methodologies in individualistic cultures will not remain true when transferred to other cultures. Consistent with this knowledge gap about the equivalence of survey modes, researchers who expect to study IT adoption behaviour involving subjects in collectivistic cultures or intend to collect data by using mixed-mode surveys to compare IT/ IS adoption behaviour in a global environment have a necessity to understand the impact of culture on MI.

The results of this study provide answers to the questions of whether, to what extent, and how national culture exerts an influence on the MI of data collected via different survey modes. There are three interesting findings. First, compared to the traditional paper-based surveys, respondents are more inclined to satisfice on Internet surveys regardless of their national cultures. In our study respondents in both the USA and China manifested significantly more satisficing behaviours on Internet surveys. In contrast to the physical contact between respondents and survey researchers in paper surveys, the Internet creates an anonymous, impersonal social situation that produces a sense of disinhibition in respondents. In addition, in the context of Internet surveys, respondents are potentially distracted by other sources such as a simultaneous Facebook conversation, listening to TV, or music in the background. On Internet surveys the respondents are also more likely to squeeze the survey in between other commitments (Barge and Gehlbach 2012). As a result of the competition for attention, fewer respondents to Internet surveys will dedicate the time and attention to the survey that researchers would prefer.

Second, the satisficing difference between paper and online surveys appears contingent on national cultures. Based on the results of our study, there is a relatively larger difference in satisficing between online and paper surveys in the samples from China that result in measurement nonequivalence. Actually, respondents in the USA manifested more satisficing behaviour in both offline and online surveys than respondents in China. However, the similar extent of satisficing in the two survey modes produced convergent results in the US sample. A potential explanation for the outcome is that people in individualistic cultures are more disposed and more willing to 'follow their own hearts' when compared with people in collectivistic cultures. People in collectivistic societies are more likely to value harmony and moderation and try to conform to the expectations of others. The

individuals in collectivistic cultures have relatively greater concern for what others think of them and stress the importance of avoiding criticism. Thus, in offline surveys, respondents perceive strong social pressure even in an anonymous offline environment, and this concern compels them to expend adequate efforts to participate in the survey. However, in Internet surveys, respondents may still abide by the requests of survey researchers to take part in surveys because doing so is embedded in their cultural values, which include the desire not to disrupt the social norms (Tan et al. 1998). However, with the great anonymity that online offers, they probably devote relatively less time and mental effort to completing the survey versus the offline version than those in individualistic cultures. As a result, the difference gap is larger among respondents in the collectivist culture. Additionally, hedonic and social-oriented Internet consumption patterns further cause respondents to be more likely to engage in satisficing behaviour in Internet surveys.

Third, the impact of gender on satisficing appears contingent on national cultures. In individualistic cultures, males and females exhibit no significant difference in the degree of satisficing while males in collectivistic cultures are more likely to engage in satisficing in both online and offline surveys. We believe that the result is particularly related to Confucian values. In China, women are more likely to commit to social norms in comparison to men and exhibit greater conformance than men. In addition, in spite of some previous studies (e.g. Booth-Kewley et al. 2007, Barenboym et al. 2010) that speculate that males are more affected by the mode of survey administration than females, this research did not support that claim. One possible explanation is that gender exerts its effects on MI via interrelated factors such as the desire to manage impression. This is an important finding that supports the need for further work because no previous research has explored the difference in satisficing using gender as an independent variable and the findings have an impact on the conduct of research in different cultures.

There are some theoretical and practical implications of our findings. Theoretically, this study confirms the potential influence of culture on MI and investigates the underlying mechanisms. Our study shows that the collectivistic culture expands the satisficing gap between offline and online surveys, which in turn results in the potential measurement nonequivalence. On the other hand, the results also support the importance of checking data carefully even if consistent results are obtained from different survey modes. Comparable results across survey modes do not necessarily indicate accurate and effective measurements, because this study's findings show that the results are potentially contaminated by satisficing. This satisficing can also affect the scale reliabilities and the associations between the latent variables. The shortcut in cognitive processing might improve measurement reliability and validity, but in most surveys, devoting more careful thought is liable to produce more accurate replies (Holbrook et al. 2003). Practically, this study shows that, when collecting data via a mixed-mode survey approach, researchers should pay attention to the satisficing behaviour both in online surveys and in offline surveys. Examining satisficing should become routine in the data preprocessing stage. Researchers can conduct the analyses with and without satisficers present in the data sets to highlight any effect the problematic responses have on the data. Researchers can also consider weighting responses in a

manner that is inversely proportional to the degree of satisficing (Barge and Gehlbach 2012). In collectivistic cultures, researchers can consider increasing the social connection of the respondent to the researcher in surveys to alleviate satisficing behaviour on the Web. For instance, social media-based surveys represent an opportunity to augment the sociability that encourages respondents to dedicate the time and attention to surveys.

Several limitations of this work should be noted. First, our results may not generalise to other surroundings such as organisational surveys because the subjects in this study were students. Although student samples are frequently used in MI research, we hope that future research will verify our results within an organisational setting. Second, the different suitability of paper and electronic screens to display information in non-alphabetic characters (e.g. Chinese characters) may contribute to the measurement variation observed in online versus paper-based surveys.³ In contrast to English, which is an alphabetic language, Chinese is a pictographic language, which cannot be presented as clearly as alphabetic languages on electronic devices (Liao et al. 2008). However, with the emergence of electronic products that feature high-resolution displays the influence of this aspect on measurement variance across different survey modes is diminishing.

6. Conclusion

The conclusions in previous research about MI across survey modes were mainly derived from data collected in North America and without concern for possible cultural differences. Little research has been conducted to assess ME across disparate survey modes in collectivistic cultures. The present research represents an important effort to investigate whether and how national cultures affect MI. Our research shows that the relatively large satisficing discrepancy between paper and online surveys in collectivistic countries facilitates greater measurement variance among survey modes than in individualistic cultures.

Notes

1. As appropriate to the context we use the terms ME or MI to describe the equality of measurements or the inequality of measurements, respectively.
2. Similar conclusions are also reached without excluding the extreme values of 1 and 7.
3. The authors thank Dr Cakir for this point.

Acknowledgements

The authors thank the Associate Editor and the anonymous reviewers for their helpful comments. This work was supported by the National Nature Science Foundation of China (NSFC) under grant 71101018 and the Fundamental Research Funds for the Central Universities in China under grant ZYGX2010J131.

References

Abareshi, A. and Martin, B., 2009. A methodological examination of MIS survey research from 1992–2006. *Journal of Information & Knowledge Management*, 8 (2), 137–146.

Agarwal, R. and Karahanna, E., 2000. Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, 24 (4), 665–694.

Agarwal, R. and Prasad, J., 1998. A conceptual and operational of personal innovativeness in the domain of information technology. *Information Systems Research*, 9 (3), 204–215.

Bao, Y., Zhou, K., and Su, C., 2003. Face consciousness and risk aversion: do they affect consumer decision-making? *Psychology and Marketing*, 20 (8), 733–755.

Barenboym, D.A., Wurm, L.H., and Cano, A., 2010. A comparison of stimulus ratings made online and in person: gender and method effects. *Behaviour Research Methods*, 42 (1), 273–285.

Barge, S. and Gehlbach, H., 2012. Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, 53 (2), 182–200.

Behrend, T., et al., 2011. Cloud computing adoption and usage in community colleges. *Behaviour & Information Technology*, 30 (2), 231–240.

Belli, R.F., Herzog, R.A., and Van Hoewyk, J., 1999. Scale simplification of expectations for survival: cognitive ability and the quality of survey responses. *Cognitive Technology*, 4 (2), 29–38.

Booth-Kewley, S., Larson, G.E., and Miyoshi, D.K., 2007. Social desirability effects on computerized and paper-and-pencil questionnaire. *Computers in Human Behaviour*, 23 (1), 463–477.

Bosnjak, M., Tuten, T.L., and Wittmann, W.W., 2005. Unit (non)

response in web-based access panel surveys: an extended planned-behaviour approach. *Psychology & Marketing*, 22 (6), 489–505.

Braunsberger, K., Wybenga, H., and Gates, R., 2007. A comparison of reliability between telephone and web-based surveys. *Journal of Business Research*, 60 (7), 758–764.

Buchanan, T., Johnson, J.A., and Goldberg, L.R., 2005. Implementing a five-factor personality inventory for use on the internet. *European Journal of Psychological Assessment*, 21 (2), 116–128.

Calhoun, K., Teng, J., and Cheon, M., 2002. Impact of national culture on information technology usage behaviour: an exploratory study of decision making in Korea and the USA. *Behaviour & Information Technology*, 21 (4), 293–302.

Chen, F., 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14 (3), 464–504.

Clark, R. and Goldsmith, R., 2006. Interpersonal influence and consumer innovativeness. *International Journal of Consumer Studies*, 30 (1), 34–43.

Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, M., Bedeian, A., and Field, H., 2006. The measurement equivalence of web-based and paper-and-pencil measures of transformational leadership. *Organizational Research Methods*, 9 (3), 339–368.

Deutskens, E., de Ruyter, K., and Wetzels, M., 2006. An assessment of equivalence between online and mail surveys in service research. *Journal of Service Research*, 8 (4), 346–355.

Fang, J., Wen, C., and Pavur, R., 2012a. Participation willingness in web surveys: exploring effect of sponsoring corporation's and survey provider's reputation. *Cyberpsychology, Behavior, and Social Networking*, 15 (4), 195–199.

Fang, J., Wen, C., and Prybutok, V.R., 2012b. An assessment of equivalence between Internet and paper-based surveys: evidence from collectivistic cultures. *Quality and Quantity*, doi:

10.1007/s11135-012-9783-3.

Fouladi, R.T., McCarthy, C.J., and Moller, N.P., 2002. Paper-and pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment*, 9 (2), 204–215.

Fricker, S., et al., 2005. An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, 69 (3), 370–392.

Gaspay, A., Dardan, S., and Legorreta, L., 2008. Software of the mind? A review of applications if Hofstede's theory to IT research. *Journal of Information Technology Theory and Application*, 9 (3), 1–37.

Green, D. and Pearson, M., 2011. Integrating website usability with the electronic commerce acceptance model. *Behaviour & Information Technology*, 30 (2), 181–199.

Hair, J.F., et al., 1998. *Multivariate data analysis*. 6th ed. Upper Saddle River, NJ: Prentice Hall.

Heerwegh, D. and Loosveldt, G., 2007. Personalizing e-mail contacts: its influence on web survey response rate and social desirability response bias. *International Journal of Public Opinion Research*, 19 (2), 258–268.

Hofstede, G., Hofstede, G.J., and Minkov, M., 2010. *Cultures and organizations, software of the mind: intercultural cooperation and its importance for survival*. 3rd ed. New York: McGraw-Hill.

Holbrook, A., Green, M., and Krosnick, J., 2003. Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67 (1), 79–125.

Hurt, T., Joseph, K., and Cook, C., 1977. Scales for the measurement of innovativeness. *Human Communication Research*, 4 (1), 58–65.

Jarvenpaa, S. and Leidner, D., 1999. Communication and trust in global virtual teams. *Organization Science*, 10 (6), 791–815.

Kaminska, O., Mccutcheon, A., and Billiet, J., 2010. Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, 74 (5), 956–984.

Keh, H.T. and Sun, J., 2008. The complexities of perceived risk in cross-cultural services marketing. *Journal of International Marketing*, 16 (1), 120–146.

Kim, D., Ferrin, D., and Rao, H., 2009. Trust and satisfaction, two stepping stones for successful e-commerce relationships: a longitudinal exploration. *Information Systems Research*, 20 (2), 237–257.

King, W. and Miles, E., 1995. Quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: a test of measurement equivalence. *Journal of Applied Psychology*, 80 (6), 643–651.

Ko, H., Roberts, M., and Cho, C.H., 2006. Cross-cultural differences in motivations and perceived interactivity: a comparative study of American and Korean internet users. *Journal of Current Issues & Research in Advertising*, 28 (2), 93–104.

Krosnick, J.A., 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5 (3), 213–236.

Krosnick, J.A., Narayan, S., and Smith, W., 1996. Satisficing in surveys: initial evidence. In: M.T. Braverman and J.K. Slater, eds. *Advances in survey research*. San Francisco, CA: Jossey-Bass, 29–44.

La Ferle, C. and Kim, H., 2006. Cultural influences on internet motivations and communication styles: a comparison of Korean and US consumers. *International Journal of Internet Marketing and Advertising*, 3 (2), 142–157.

de Leeuw, E., 2005. To mix or not to mix? Data collection modes in surveys. *Journal of Official Statistics*, 21 (2), 1–23.

Leidner, D.E. and Kayworth, T., 2006. A review of culture in information systems research: toward a theory of information technology culture conflict. *MIS Quarterly*, 30 (2), 357–399.

Levine, R., Norenzayan, A., and Philbrick, K., 2001. Crosscultural

differences in helping strangers. *Journal of Cross-Cultural Psychology*, 32 (5), 543–560.

Liao, H., Proctor, R.W., and Salvendy, G., 2008. Content preparation for cross-cultural e-commerce: a review and a model. *Behaviour & Information Technology*, 27 (1), 43–61.

Lu, J., Yao, J.E., and Yu, C.S., 2005. Personal innovativeness, social influences and adoption of wireless internet services via mobile technology. *Journal of Strategic Information Systems*, 14 (3), 245–268.

Meade, A.W., Michels, L.C., and Lautenschlager, G.J., 2007. Are internet and paper-and pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods*, 10 (2), 322–345.

Miles, E.W. and King, W.C., 1998. Gender and administration mode effects when pencil-and-paper personality tests are computerized. *Educational and Psychological Measurement*, 58 (1), 68–76.

Nunnally, J.C. and Bernstein, I.H., 1994. *Psychometric theory*. 3rd ed. New York: McGraw-Hill.

Raju, S., Laffitte, J., and Byrne, M., 2002. Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87 (3), 517–529.

Roberts, C., 2007. Mixing modes of data collection in surveys: a methodological review. NCRM review papers 008. Southampton: ESRC National Centre for Research Methods.

Satorra, A. and Bentler, P.M., 2001. A scaled difference-chi square test statistic for moment structure analysis. *Psychometrika*, 66 (4), 507–514.

Setterstrom, A., Pearson, M., and Orwig, R., 2012. Web-enabled wireless technology: an exploratory study of adoption and continued use intentions. *Behaviour & Information Technology*, doi: 10.1080/0144929X.2012.708785.

Stephenson, L.B. and Cr. te, J., 2011. Studying political behaviour: a comparison of internet and telephone surveys. *International Journal of Public Opinion Research*, 23 (1), 24–55.

Tan, B., et al., 1998. Computer-mediated communication and majority influence: assessing the impact in an individualistic and a collectivistic culture. *Management Science*, 44 (9), 1263–1278.

Tourangeau, R., Rips, L., and Rasinski, K., 2000. *The psychology of survey response*. Cambridge: Cambridge University Press.

Triandis, H.C., 1988. Collectivism vs. individualism: a reconceptualization of a basic concept in cross-cultural psychology. In: G. Verma and C. Bagley, eds. *Cross-cultural studies of personality, attitudes and cognition*. London: Macmillan, 60–95.

Vandenberg, R.J., 2002. Toward a further understanding of and improvement in measurement invariance methods and procedure. *Organizational Research Methods*, 5 (2), 139–158.

Vandenberg, J. and Lance, E., 2000. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3 (1), 4–70.

Westland, J.C., 2010. Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications*, 9 (6), 476–487.

Zhang, N., et al., 2009. Impact of perceived fit on e-government user evaluation: a study with a Chinese cultural context. *Journal of Global Information Management*, 17 (1), 49–69.

Zhou, T., 2011. An empirical examination of users' post-adoption behaviour of mobile services. *Behaviour & Information Technology*, 30 (2), 241–250.