

## Eastern Illinois University The Keep

---

Faculty Research & Creative Activity

Business, School of

---

September 2012

# An assessment of equivalence between Internet and paper-based surveys: evidence from collectivistic cultures

Jiaming Fang

*University of Electronic Science and Technology of China*

Chao Wen

*Eastern Illinois University, [cwen@eiu.edu](mailto:cwen@eiu.edu)*

Victor R. Prybutok

*University of North Texas*

Follow this and additional works at: [http://thekeep.eiu.edu/business\\_fac](http://thekeep.eiu.edu/business_fac)



Part of the [Business Commons](#)

---

### Recommended Citation

Fang, Jiaming; Wen, Chao; and Prybutok, Victor R., "An assessment of equivalence between Internet and paper-based surveys: evidence from collectivistic cultures" (2012). *Faculty Research & Creative Activity*. 6.

[http://thekeep.eiu.edu/business\\_fac/6](http://thekeep.eiu.edu/business_fac/6)

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](mailto:tabruns@eiu.edu).

# **An assessment of equivalence between Internet and paper-based surveys: evidence from collectivistic cultures**

Jiaming Fang, Chao Wen, and Victor R. Prybutok

**Abstract:** Little research exists that addresses the equivalence in collectivistic cultures of paper- versus Internet-based surveys. This study addressed this gap and examined the measurement equivalence of individual innovativeness scales between Internet surveys and paper-based surveys within a collectivistic culture (with China serving as our example). The study analyzed and compared survey data from both paper and web-based surveys using confirmatory factor analysis. The assessment of invariance included the levels of configural, metric, scalar, and covariance invariance. The means and variance of latent variables were also compared. The results show that measurements are invariant at the two levels (configural and metric), and the covariances between latent variables are also equivalent, but the mean and variance differences of latent variables are apparent. The results indicate that when conducting research in collectivistic cultures and collecting data from distinct survey modes, researchers should concern themselves with the potential of extreme response patterns and the inclination of social desirability responding, as well as considering the measurement invariance across survey modes.

## **1 Introduction**

The use of the Internet to collect data is restrained by coverage limitations and the lack of willingness of people to respond. Numerous studies were conducted on approaches for promoting the willingness of potential respondents to participate in Internet surveys, and researchers have documented numerous reasons why participants don't respond to Internet surveys (e.g. Bosnjak et al. 2005; Fan and Yan 2010; Fang et al. 2012; Göritz 2006; Heerwegh and Loosveldt 2007). However, much less research has addressed the problem of coverage limitation. Low coverage rates cause potential inaccuracies in the questionnaire study results.

One solution proposed to alleviate the negative effect of a low coverage rate is to use a mixed-mode design. For example, using an Internet survey to collect data from respondents who can access the Internet and then accessing other respondents by collecting data via another delivery mode such as a paper-based survey.

When integrating data collected from Internet surveys with traditional paper-based surveys, researchers and practitioners must ensure the measurement equivalence (ME) of these two modes and that the measured latent construct has the same theoretical pattern under different survey modes (Miles and King 1998). If this invariance assumption is violated, it is not appropriate to combine the data gathered

by distinct survey administration modes.

Numerous studies were conducted to examine the equivalence of Internet surveys with their paper-based counterparts (e.g. King and Miles 1995; Booth-Kewley et al. 2007; Meade et al. 2007), but almost all of this research was conducted in individualistic cultures, little empirical research investigated whether paper and online surveys produced convergent results in collectivistic cultures. In addition, most existing MI studies are conducted in such research areas as psychology or public opinion. These studies (e.g., Vecchione et al. 2012) mainly focus on the scales measuring noncognitive skills such as personality, depression, and IQ. Few studies have assessed MI of attitude and self-evaluation scales which are widely used in behavior research.

To fill that gap in the literature, we employed multigroup confirmatory factor analysis to examine MI of the individual innovativeness scale. Innovativeness is one of the most important constructs related to the study of consumer behavior (Tellis and Chandrasekaran 2010), thus the individual innovativeness scale is one of the most frequently used scales in consumer behavior, information systems/information technology (IS/IT) adoption, and E-commerce research. We use data collected by paper-based surveys, and two Internet-based survey modes (using an online forum and a website) in China, a typical collectivistic culture. We expect the difference in perception of anonymity in these surveys will lead to a lack of ME because people in a collectivistic culture are particularly concerned about how they appear to others. In the offline environment, individuals in collectivistic cultures are more sensitive to the evaluations of the others than their counterparts in individualistic cultures. Collectivists desire to enhance, maintain, and avoid losing face in relation to other people in social activities. Individuals stress their social needs more than their private needs when making decisions. Thus, in offline contexts, respondents are inclined to engage in socially desirable responding (SDR), however, in the online anonymous environment, the reduced social pressure potentially increases the respondents' candor.

This study contributes to the extant literature on response biases, MI assessments and cross-nation survey research by offering empirical evidence of MI testing between Internet and paper-based surveys in collectivistic cultures. The results provide researchers, those in collectivistic cultures and those who need to collect data in collectivistic cultures, information on the feasibility of integrating data collected via Internet surveys with the paper-based surveys. Our research also further extends the work of Davison et al. (2006), who investigated the feasibility of collecting data using the Internet as a medium in collectivistic countries by exploring the likelihood to integrate the data from web surveys with paper-based surveys.

The remainder of this paper is organized as follows. First, we review the literature on MI between online and paper-based surveys and identify the potential mechanism of collectivistic cultures' impact on measurement inequality. Second, we

present the research methodology and results of conducting the data analysis. Finally, we conclude with the theoretical and practical implications of our findings, their limitations and recommendations for further research.

## **2 Background**

### **2.1 MI between online- and paper-based survey modes**

MI is the extent to which measurements conducted under different conditions measure the same attributes (Miles and King 1998). To integrate the data collected from Internet surveys with data from paper-based surveys, one must assume that the data gathered by these two survey modes are comparable. That is, one must assume that the survey has MI across these two different survey modes and that the numerical values under consideration are measured on the same scale (Rosen 2004). A failure to demonstrate ME across survey modes not only casts doubt on their comparability but also may make unambiguous interpretation impossible (Vandenberg and Lance 2002; Cole et al. 2006).

Some prior studies have shown that there was no evidence of administration mode affecting ME (e.g. Davidov and Depner 2011; King and Miles 1995; Leidner and Kayworth 2006), other studies indicated that survey modes affected the results (e.g. Fouladi et al. 2002; Buchanan et al. 2005). Still other research found that Internet surveys and paper-based surveys possessed partial MI (e.g. Meade et al. 2007; Steinmetz et al. 2009). Although the results remain inconclusive, these studies have greatly increased our understanding of MI between paper and online surveys. The lack of consistency presents an important area of research in identifying the conditions that allow for MI versus measurement variance.

It is noteworthy that although a number of empirical studies were conducted to examine the differences in observed scores in online surveys and paper-based surveys, almost all of these MI studies to date have been carried out in the same social-cultural backgrounds, that is, individualistic cultures. 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. Furthermore, the existing MI studies widely use single factor model; however, behavior research scholars are likely to propose models that involve multiple latent factors and focus on the relations between latent variables. When collecting data to test a model with more than one underlying factor, the equivalence of the theoretical structure, especially the factor covariances, is of interest (Reise et al. 1993; Vandenberg 2002). Hence, it is valuable and essential to examine the invariance of the measurement model possessing two or more latent variables rather than the one factor model tested in the prior behavioral research.

In addition, Internet surveys are distributed via four major channels: E-mail, newsgroup, online forums, and survey websites (Chou 1997). However, previous studies refer to Internet surveys based on the use of an Internet survey website; for

example, see Yetter and Capaccioli (2010). However, survey models are different when serving as a platform rather than a website URL delivering tools. In particular, the anonymity of the responses varies. For example, if the online forum acts as a survey platform, the respondents can see the replies of the previous respondents while, whilst in website surveys the results are invisible to the other respondents. Surveys administered in an anonymous environment potentially create a social situation that produces a sense of disinhibition in respondents, and this sense of disinhibition increases the likelihood of more candid reports. Booth-Kewley et al. (2007) showed that respondents had higher self-deceptive enhancement (SDE) scores than anonymous respondents. Therefore, it is necessary to examine other forms of the Internet surveys to gain a better understanding of how the format impacts MI.

## 2.2 Collectivistic cultures impact on measurement inequality between online and paper surveys

A compelling explanation of potential sources of MI is that Internet surveys create the unique impersonal social situation in which individuals feel more anonymous, less inhibited, less concerned about how they appear to others, and thus respondents are less inclined to engage in SDR in Internet surveys (Booth-Kewley et al. 2007). SDR is the inclination to give answers that make the respondent appear good, or the propensity to make a good impression. Respondents experience the desire to achieve greater social desirability in an environment where they are identified rather than anonymous (Paulhus 1984). SDR is a particularly prominent concern in interviewer-administered modes such as face-to-face or telephone surveys. Researchers speculate that respondents may be less prone to give socially desirable answers on the computer and may respond with a greater degree of candor for an impersonal social situation created by computers (Booth-Kewley et al. 2007).

Prior research has identified two types of socially desirable responding: impression management (IM) and SDE. IM is the deliberate propensity to overreport desirable behaviors and underreport undesirable ones, while SDE is the tendency to give honestly believed but actually over positive reports about oneself (Booth-Kewley et al. 2007). A meta-analysis conducted by Dwight and Feigelson (2000) to determine the extent to which SDR were provided by computerized testing and the results showed a small but statistically significant effect for IM, with lower IM assessment in the computer condition. A recent study (Lalwani and Chiu 2008) further suggests that collectivists are especially likely to provide normatively desirable and socially appropriate responses and, hence, to engage in IM, whilst individualists have a propensity to engage in SDE.

It is speculated that the MI of data collected via different survey modes in collectivistic cultures is more likely to suffer because people in collectivistic cultures are more sensitive to the evaluations of the others than their counterparts in individualistic cultures (Higgins et al. 2002). This sensitivity is especially prevalent

in situations involving attitude expressions and self-assessments. Individualism emphasizes personal freedom and responsibility; collectivism endorses social relatedness and interdependence with others in one's family or social group. Compared to individualists, people in collectivistic cultures are more concerned about face consciousness which refers to people's desire to enhance, maintain, and avoid losing face in relation to other people in social activities (Bao et al. 2003). Face consciousness plays an important role in collectivistic countries, and it drives collectivists to seek social approval and avoid social disapproval (Keh and Sun 2008). Individuals place more emphasis on publicly visible social-self rather than express their internal selves so as to maintain or enhance their faces. The result is that respondents in collectivistic culture are more likely to self-disclose in Internet surveys than in paper-based surveys. The apparently disparate levels of SDR involved in different survey modalities lead to divergent results across the survey modes.

### **3 Methodology**

#### **3.1 Scale selection**

We selected two scales to examine MI across the two survey modes. One is the personal innovativeness in the domain of IT (PIIT) as proposed by Agarwal and Parsad (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 research such as consumer innovativeness management, IS/IT management. PIIT scale was used worldwide with the evidence consistently showing that PIIT is an important predictor of individual IS/IT adoption behavior. The second scale is global innovativeness (GI), which measures an individual's general predisposition 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 IS/IT applications. Given that PIIT belongs to the domain of a specific innovativeness scale, we posit a positive correlation between PIIT and GI.

These two scales were selected based on the following considerations. First, PIIT scale was initially developed for use with paper surveys in individualistic cultures, while PIIT is now increasingly assessed using web-based methods. Researchers have seemingly assumed that paper and web-based surveys exhibit adequate cross-mode equivalence. Studies have yet to examine whether administration mode has an effect on the MI of the PIIT scale. We suspected that respondents in collectivistic cultures would provide more self-disclosure in Internet surveys than in online forum or paper-based surveys because of the difference in perceived anonymity in the different modes. Therefore, it is worthwhile to assess the MI of innovativeness scales across different survey modes. Second, although a number of studies examined MI based on a single factor CFA model, models with more than one latent factor are commonly used in the field of behavior research. Technically, a multi-factor CFA model enables us to assess the equality of  $\Lambda$  matrices (i.e. covariance

matrix among the underlying factors across survey modes). Third, many MI research has focused on the scales measuring noncognitive skills; however, substantive research involves cognitive measurements such as attitude and self-perception assessments in the behavioral research domain. Thus, substantive research is still needed in the field of behavior to engage cognitive measures. The innovativeness scales belong to this domain. Finally, both scales are reasonably parsimonious. Extensive redundancy in items puts a burden on respondents, which can potentially incur extra confounding effects on the findings. The scales have simple structures, the same question formats (a seven-point Likert-type response format), and the survey questions themselves were pertinent to the student population, so respondents can easily understand both scales in a single instrument.

### 3.2 Data collection

We administered the surveys that included the innovativeness scales via three survey modes: paper-based survey, online forum (message boards)-based Internet survey, and website-based Internet survey. Internet surveys are distributed via four major channels. Because of Internet worms, distinct terminal operations systems, and text-reading software, E-mail is now primarily used to send survey invitations to potential respondents rather than as a carrier of survey questionnaires. Furthermore, surveys are distributed through news systems in ways similar to the online forum approach (Chou 1997). Consequently, online forum-based surveys and website-based surveys encompass most of the same characteristics as Internet surveys.

Our MI assessments were based on Chinese samples. China is considered a highly collectivistic culture, is rapidly advancing technologically, and has the largest population in the world. Therefore, we decide to use China as the source for data in a collectivistic culture. Students from a large public university in the southwest of China were recruited to participate in the paper-based surveys. We distributed 300 questionnaires in classes, and respondents completed the questionnaires in their classroom settings as part of the class activity. Students submitted the completed anonymous questionnaires to researchers that stated they would keep the files confidential. The final sample included 248 complete and valid questionnaires. 40% male and 60% female respondents were, in this sample, within an age range of 18–36 years ( $M = 20.56$ ,  $SD = 3.14$ ).

Forum-based Internet surveys were conducted using a campus online forum at the same university. Students have to register with the forum and then subsequently log in in order to post messages. The forum was characterized by the real name registration, and each student has only one account. Users can specify any avatars or usernames to represent themselves on the forum if they wish. Each account could only participate in a survey once and none of the classes that had the opportunity to take the paper survey were given access to the Internet survey. We obtained 228 valid responses. 84% male and 16% female respondents were, in this sample,

within an age range of 19–32 years ( $M = 21.40$ ,  $SD = 1.75$ ).

Website-based Internet surveys were conducted using a well-established Internet survey platform ([www.zhijizhibi.com](http://www.zhijizhibi.com)) in China. This platform is similar to SurveyMonkey.com or Qualtrics.com in the US. E-mail addresses for the potential respondents were obtained in advance and the potential student respondents were informed in classes that the survey URL would be sent using E-mail invitations. In 2 weeks we received 263 questionnaires. Of those we received 45% were male and 55% were female respondents, within an age range of 18–34 years ( $M = 20.18$ ,  $SD = 1.71$ ).

To evaluate the demographic similarity among the three groups, we made comparisons on age, gender, major (social science or nature science) and web experience. A Kruskal–Wallis test (age) and a series of  $\chi^2$  tests (gender, major, and experience) showed that there were no statistically significant differences on those variables between the respondents in the website- and paper-based group. However, the composition of gender was significantly different between the online forum group and the other two groups ( $\chi^2$  (with  $df = 2$ ) = 108.5,  $p < 0.001$ ).

Since demographic variables are commonly used as proxy variables for cognitive ability, the similar demographic variables indicated that samples in this research possessed similar cognitive ability.

The use of student participants is potentially questioned on grounds of external validity (Mcknight et al. 2002). However, in this study the use of school students did not present a significant threat to validity for several reasons. First, the students' profiles and performance were similar to the target population. According to the Chinese Network Information Report of 2012, Internet users below age 29 account for about 60% of all Internet users in China (530 million). These young Internet surfers are also the predominant participants in Internet surveys. Second, because our research required examination of MI across distinct survey modes, we can assess MI while controlling the demographic of respondents if we recruit homogeneous groups. Since prior research showed that SDR was related to important demographic characteristics of respondents such as age, gender, education level and income (Davidov and Depner 2011; Steinmetz et al. 2009), by controlling for the demographic variables, we could exclude the interference effects of these variables.

To verify the adequacy of the sample size to conduct MCFA, this investigation adopted the suggestions of Westland (2010) to calculate the ratio of indicators to latent variables. Results showed that valid samples in the survey modes were adequate to use CFA analysis.

### 3.3 Testing procedures

We employed multigroup confirmatory factor analysis to assess MI between



Internet and paper survey modes. The testing procedures involved comparing a series of increasingly stringent models by sequentially constraining different parameter estimates to be invariant across survey modes (French and Finch 2008).

Consistent with prior research (Vandenberg and Lance 2002; Vandenberg 2002), we examined the equality of the observed variance-covariance matrices first. Then 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 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 the equality of  $\gamma$  matrices to assess if metric invariance was maintained. Finally, we evaluated the equality of item intercepts (scalar invariance) and the invariance of the covariance matrix among the factors across the survey modes.

The  $\chi^2$  goodness of fit and three index measures (Tucker–Lewis Index [TLI], Normed Fit Index [NFI], and root mean square error of approximation [RMSEA]) provided the general model fit measurement 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 2002; Leidner and Kayworth 2006; Chen 2007). Although the  $\chi^2$  difference statistic has been the primary criterion, the statistic is sensitive to sample size and violation of the multivariate normality assumption (Chen 2007). As rules of thumb, TLI and NFI 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  $\alpha$  of GI and PIIT was 0.85 and 0.75, respectively. The corresponding  $\alpha$  coefficient was 0.72 and 0.73 in the online forum-based survey group, and 0.82 and 0.75 in the websitebased survey group. These reliability coefficients were all above the suggested cutoff value 0.7 (Nunnally and Bernstein 1994).

In Table 1 below, means and standard deviations of observed items across survey modes are listed. We tested the equality of variances of each observed item across distinct survey modes using Levene's robust test. Table 1 also shows that the variances were not equal. To arrive with this result, we conducted Kruskal–Wallis tests to inspect mean differences of observed items, and the results indicated that significant differences existed.

### 4.2 MI assessment

#### 4.2.1 Test of variance–covariance matrices invariance

We evaluated equality of covariance matrices across three survey modes through the  $\chi^2$  statistic, and the results showed that  $F = 5.82$  ( $p < 0.001$ ),  $\chi^2(56) = 326.06$  ( $p < 0.001$ ). The results rejected the null hypothesis, so it was necessary to identify the source of nonequivalence (Vandenberg and Lance 2002).

**Table 1** Means and standard deviations of observed items across survey modes

Observed items	Paper-based		Forum-based		Website-based		Group difference	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Mean	Variance
GI1	5.81	1.39	5.16	1.34	5.18	1.21	$p < 0.01$	$p < 0.05$
GI2	5.79	1.36	5.30	1.33	5.11	1.13	$p < 0.01$	$p < 0.01$
GI3	5.17	1.36	4.75	1.33	4.80	1.09	$p < 0.01$	$p < 0.01$
PIIT1	5.35	1.53	5.17	1.59	5.02	1.08	$p < 0.01$	$p < 0.01$
PIIT2	4.07	1.82	4.50	1.51	4.27	1.18	$p < 0.05$	$p < 0.01$
PIIT3	3.81	1.60	4.23	1.38	4.27	1.14	$p < 0.01$	$p < 0.01$
PIIT4	4.15	1.68	4.71	1.64	4.07	1.01	$p < 0.01$	$p < 0.01$
Mean difference	GI: $p < 0.01$				PIIT: $p < 0.01$			
Variance difference	GI: $p < 0.01$				PIIT: $p < 0.01$			

*GI<sub>i</sub>* and *PIIT<sub>i</sub>* denote the *i*th measurement item of the GI and PIIT scales, respectively

#### 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 showed that seven items loaded onto two factors (GI and PIIT) and cross-loadings were well below the usual cutoff value of 0.40. Further, we tested whether the configural invariance held across survey modes necessary. We constrained the means of two latent variables in the two survey modes to zero to circumvent the identification problem (Bentler 2005). The results are listed in Table 2.

The fit indices indicated that we could not reject the hypothesis that the equality of the number of factors and factor pattern matrices were held in each survey mode. This model (M1) provided the baseline model value against which subsequent tests for invariance could be compared.

#### 4.2.3 Test of invariance of factor loadings matrix

Because configural invariance had been established, we further evaluated metric invariance across survey modes (M2). The results are shown in Table 3.

In the above illustration, after constraining the factor loadings to invariance across

survey modes, the nested model was still well-fitting, and the change of  $\chi^2$  was not significant ( $p = 0.19$ ). Therefore, metric invariance is established. Accordingly, we could proceed in testing for the invariance among the factors.

**Table 2** Test of configural invariance

Model	$\chi^2$	<i>df</i>	TLI	NFI	RMSEA
M1	100.75	39	0.94	0.94	0.05

Model M1 constrains the number of factors and factor pattern to invariance across the survey modes *TLI* Tucker–Lewis Index, *NFI* Normed Fit Index, *RMSEA* root mean square error of approximation

**Table 3** Result of invariance of factor loadings matrix assessment

Model	$\chi^2$	<i>df</i>	TLI	NFI	RMSEA	$\Delta\chi^2$	$\Delta df$	<i>p</i>
M2	114.34	49	0.95	0.93	0.04	13.58	10	0.19

Model M2 constrains the factor loadings to invariance across the survey modes *TLI* Tucker–Lewis Index, *NFI* Normed Fit Index, *RMSEA* root mean square error of approximation

**Table 4** Results of invariance of item intercept assessment

Model	$\chi^2$	<i>df</i>	TLI	NFI	RMSEA	$\Delta\chi^2$	$\Delta df$	<i>p</i>
M3	189.87	61	0.92	0.89	0.05	71.50	10	<0.001

Model M3 constrains the factor loadings, the covariances, the variances and item intercepts to invariance across the survey modes *TLI* Tucker–Lewis Index, *NFI* Normed Fit Index, *RMSEA* root mean square error of approximation

#### 4.2.4 Test of invariance of item intercepts across survey modes

To further verify 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). Model M3 was nested within Model M2. As shown in Table 4, a significant  $\chi^2$  change along with the NFI value below 0.90 indicated that the intercept equality assumption was not supported.

#### 4.2.5 Test of invariance of covariance matrix among the factors

We adopted a two-step testing method to assess the invariance of the covariance matrix. We first constrained the covariance to equality based on Model M3 while keeping the variances of factors freely estimated (Model M4). Since scalar invariance was not verified, we constructed Model M4 based on the M2. Obviously, M4 was nested within Model M2, so we could assess the model fit compared to the Model M2.

This comparison yielded a  $\chi^2$  difference value of 4.03 with 2 degree of freedom (d f ), which was statistically insignificant ( $p = 0.13$ ). The results are shown in Table 5.

Because ModelM4 was not rejected, the variances of respective factors were further fixed to invariance across surveymodes based on ModelM4 (ModelM5) and the results were listed in Table 6. Table 6 shows a significant  $\chi^2$  change which indicates that the variance equality assumption was not supported. The Akaike information criterion (AIC) value also substantiated the results. AIC=184.37 for the covariance equality model, while AIC=197.61 for the variance-covariance invariance model.

**Table 5** Result of invariance of covariance matrix assessment

Model	$\chi^2$	<i>df</i>	TLI	NFI	RMSEA	$\Delta\chi^2$	$\Delta df$	<i>p</i>
M4	118.37	51	0.95	0.93	0.04	4.03	2	0.13

Model M4 constrains the factor loadings and the covariances to invariance across the survey modes  
*TLI* Tucker–Lewis Index, *NFI* Normed Fit Index, *RMSEA* root mean square error of approximation

**Table 6** Result of invariance of covariance matrix assessment after fixing variances invariance

Model	$\chi^2$	<i>df</i>	TLI	NFI	RMSEA	$\Delta\chi^2$	$\Delta df$	<i>p</i>
M5	135.60	53	0.95	0.92	0.05	17.23	2	<0.001

Model M5 constrains the factor loadings, the covariances and the variances to invariance across the survey modes  
*TLI* Tucker–Lewis Index, *NFI* Normed Fit Index, *RMSEA* root mean square error of approximation

**Table 7** Mean and variance of latent factors across survey modes

	Online forum group ( <i>n</i> = 228)				Website group ( <i>n</i> = 255)			
	<i>b</i>	SE	CR	<i>p</i>	<i>b</i>	SE	CR	<i>p</i>
GI mean	-0.52	0.10	-5.02	<0.01	-0.50	0.09	-5.30	<0.01
Var(GI)	0.79	0.12	6.88	<0.01	0.72	0.09	7.78	<0.01
PIIT mean	-0.17	0.13	-1.29	0.20	-0.30	0.11	-2.72	<0.01
Var(PIIT)	0.91	0.15	6.08	<0.01	0.49	0.08	6.53	<0.01

Paper-based survey mode is selected as the reference group (latent variable means and variances are specified as 0 and 1, respectively). The means and variances shown in the table are calculated compared to the reference group

Model M5 indicated that the variance of latent factors was relatively large and the subsequent partial MI tests also showed that partial MIs were nonexistent.

#### 4.3 Estimating mean and variance differences on latent factors

According to Reise et al. (1993) there were several ways to evaluate the magnitude of the differences among the groups about the mean and variance of the latent

factors. In this study, we adopted the method of contrasting parameter estimates across survey modes relative to their standard errors. As mentioned previously, we standardized the factors in the reference group (paper-based survey mode). Thus, we could estimate the mean and variance relative to the zero mean and unit variance in the reference group. The results are listed in Table 7, which demonstrated that all but the mean of PIIT in the online forum survey mode were significantly smaller than those in the paper-based survey mode. The variance of GI in the online forum group (95% CI: 0.66–0.96), variance of GI (95% CI: 0.61–0.86) and PIIT (95% CI: 0.42–0.59) in the website group were significantly smaller than the corresponding variance in the paper-based group. However, the 95% CI of variance of PIIT in the online forum group (95%CI: 0.76–1.10) contains the value 1.0. Consistent with 1 in the confidence interval, we cannot reject the null hypothesis of equal variance of PIIT between the online forum group and the paper-based group.

#### 4.4 Robustness checks

The proportion of males to females varied between the online forum group and the other two groups. Previous studies indicated that women were more likely to have higher innovativeness than men (Ha and Stoel 2004), and that males might be more affected by survey administration mode than females (Booth-Kewley et al. 2007). Therefore, it was necessary to verify whether the research findings were caused by the different gender composition rather than the survey modes.

**Table 8** Impacts of gender on latent variables mean within a survey mode

	Paper-based group		Online forum group		Website group	
	PIIT	GI	PIIT	GI	PIIT	GI
$t( z )$	3.22*	1.14	0.88*	0.16	1.61	2.17
$p$ (two-tailed)	<0.01	0.26	0.38	0.87	0.11	0.03

$t(|z|)$  denotes the results of group mean-comparison tests (groups are split by gender variable). According to variance-comparison tests and normality tests, the items labeling "\*" used  $t$ -test ( $t$ ), others test employed Wilcoxon rank-sum test ( $|z|$ )

**Table 9** Impacts of gender on latent variable mean across survey modes

	PIIT		GI	
	Female	Male	Female	Male
$\chi^2(2)$	3.77	3.45	29.29	27.68
Significance	0.15	0.18	<0.01	<0.01

$\chi^2(2)$  denotes the results of Kruskal–Wallis rank tests (groups are split by the three survey modes).  $\chi^2$  Statistics have been corrected for ties

We examined the impact of gender on factors within and across survey modes, the results of which are listed in Tables 8, 9, respectively. As reported in Table 8, gender exerts a significant impact on PIIT in the paper-based group and on GI in the website group (male was higher than female). It is worth noting that although a difference exists in GI in the website group, evidence from the statistical analysis showed that the probability that a random drawing from the first population (male) was a larger value than a random value drawn from the second population (female) was only about 57.8%.

Table 9 shows that gender does not have an impact on the mean of PIIT across survey modes. However, participants offered varying answers across survey modes on GI. Post hoc multiple comparisons indicated that males filled in significantly higher scores ( $p < 0.01$ ) on GI in the paper-based survey (5.6 vs. 5.1, 5.0), while in the other two survey modes similar answers were provided. Females provided higher scores in paper surveys compared to those in website surveys, however, there is no significantly different score discrepancies between paper surveys and online forum surveys. Generally, for gender, there is no systematic influences on means of latent variables exist. The differences in the latent variable mean were not caused by gender, and could only be attributed to using disparate survey modes.

## **5 Discussion**

An increasing proportion of empirical research has been conducted using Internet surveys to collect data. Mixed-mode survey methods can be used to collect data to increase response rates, reduce survey costs and alleviate the negative effects of a low coverage rate, which may be caused by a single mode survey design. However, it is feasible to use the Internet as a component of mixed-mode surveys only when the underlying construct being measured has the same theoretical structure for each survey mode.

It is also worth noting that almost all MI studies to date were carried out in individualistic cultures and that there was little empirical evidence coming from a collectivistic culture. This is an important area of research because 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. Prior research indicated that a collectivistic culture and Confucian values could impact behavior patterns (Leidner and Kayworth 2006; Zhang et al. 2009). Furthermore, the need to capture data via multiple methods is greater in countries with a large digital divide and the equivalence of the data becomes more important if the intent is to combine data from multiple sources. Therefore, it is imperative to examine the MI between Internet- and paper-based surveys in collectivistic cultures.

The chief objective of this paper is to investigate the MI between Internet surveys and traditional paper surveys in collectivistic cultures using China as an example. We used a prevalent innovativeness scale in the consumer behavior research domain as an example and the results indicated that although the data collected

from different survey modes can obtain an equal theoretical structure (the same covariance between latent variables), the impersonal online environment engenders different latent variable mean scores and incurs more extreme responses.

Although the assumption of equality of item intercepts across groups or modes was rejected, it was not beyond our expectations. In contrast to paper-based surveys and an online forum survey, a web-based survey creates an anonymous, impersonal social situation that produces a sense of disinhibition in respondents. The individuals in collectivistic cultures emphasize concern for what others think of them and stress the importance of avoiding criticism. People in collectivistic cultures are more inclined to engage in SDR, especially in IM, than their counterparts in individualistic cultures. In an anonymous environment, individuals exhibit more self-disclosure. Thus, we expected latent mean differences across survey modes. Compared to two other groups, the website group exhibits lower mean scores. Although our paper-based survey was conducted anonymously in a classroom setting, the surroundings may still affect the respondents because they are so set in their cultural values and the perceived distance between the respondent and the researcher is relatively small.

Another interesting result is that the variances in responses obtained by the Internet surveys are much larger than those acquired from the paper-based survey. Prior research conducted in individualistic cultures showed that respondents reported more extreme responses in computerized measurements (King and Miles 1995). The phenomenon of extreme responses also exists in collectivistic cultures. Cross-cultural work showed that compared with people in individualistic cultures, people in collectivist, group-oriented societies were more likely to value harmony, face-saving, modesty, and moderation; to use the midpoint of a rating scale; and to give fewer extreme responses in paper-based survey (Higgins et al. 2002). However, in an online environment, the social anonymity is likely to lead to the emergence of extreme responses.

In addition, although some prior studies (e.g. Booth-Kewley et al. 2007; Barenbooy et al. 2010) indicated that males might be more affected by the mode of survey administration than females, this research did not support that claim. Our result is consistent with Chiu et al. (2005), Miles and King (1998) and Rosen (2004). Since few studies exploring the effects of computerized self-report measures have assessed gender as an independent variable, we expect that future research can further examine this topic.

Several limitations of this work should be noted. First, due to the use of students as a research sample, our results may not generalize to other surroundings, such as organizational surveys. Although student samples are frequently adopted in MI research, we hope that future research can verify our results in organizational settings. Second, in this study we examined the MI using CFA and did not involve structural invariance or relational equivalence assessment except for covariance-invariance. Therefore, we recommend that future research



assess the structure invariance in addition to MI using the SEM-based methods.

## **6 Conclusion**

The present research indicates that, in addition to carefully designed and psychometrically sound scales, researchers should also pay close attention to the MI when they collect data through Internet surveys or integrate Internet survey data with other survey modes in collectivistic countries such as China, Malaysia, and Japan. Different scales may demonstrate varied levels of MI. We believe that if the survey questions are more relevant to cognitive skill measures, then the MI assessment is much more important.

Furthermore, although it is difficult to obtain random samples in an online forum survey mode, our study indicates that it is still feasible to choose online forums to distribute the questionnaires under some conditions. Online forums aggregate individuals having similar interests to form a particular community. Therefore, when a survey refers to a specific topic such as the factors that influence a smart phone purchase, it is reasonable to conduct a survey in an online forum.

**Acknowledgements** This work was supported by the National Nature Science Foundation of China under Grant 71101018 and the Fundamental Research Funds for the Central Universities under Grant ZYGX2010J131.

## **References**

- Agarwal, R., Karahanna, E.: Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. *MIS Q.* 24(4), 665–694 (2000)
- Agarwal, R., Parsad, J.: A conceptual and operational of personal innovativeness in the domain of information technology. *Inf. Syst. Res.* 9(3), 204–215 (1998)
- Bao, Y., Zhou, K., Su, C.: Face consciousness and risk aversion: do they affect consumer decision-making?. *Psychol. Mark.* 20((8), 733–755 (2003)
- Barenboym, D.A., Wurm, L.H., Cano, A.: A comparison of stimulus ratings made online and in person: gender and method effects. *Behav. Res. Method* 42(1), 273–285 (2010)
- Bentler, P.M.: *EQS 6 Structural Equations Program Manual*. Multivariate Software, Encino (2005)
- Booth-Kewley, S., Larson, G.E., Miyoshi, D.K.: Social desirability effects on computerized and paper-and-pencil questionnaire. *Comput. Hum. Behav.* 23(1), 463–477 (2007)
- Bosnjak, M., Tuten, T.L., Wittmann, W.W.: Unit (non) response in web-based access



panel surveys: an extended planned-behavior approach. *Psychol. Mark.* 22(6), 489–505 (2005)

Buchanan, T., Johnson, J.A., Goldberg, L.R.: Implementing a five-factor personality inventory for use on the Internet. *Eur. J. Psychol. Assess.* 21(2), 116–128 (2005)

Chen, F.: Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Model.* 14(3), 464–504 (2007)

Chiu, Y., Lin, C., Tang, L.: Gender differs: assessing a model of online purchase intentions in E-tail service. *Int. J. Serv. Ind. Manag.* 16(5), 416–435 (2005)

Chou, C.: Computer networks in communication survey research. *IEEE Trans. Prof. Commun.* 40(3), 196–208 (1997)

Clark, R., Goldsmith, R.: Interpersonal influence and consumer innovativeness. *Int. J. Consum. Stud.* 30(1), 34–43 (2006)

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

Davidov, E., Depner, F.: Testing for measurement equivalence of human values across online and paper-and-pencil surveys. *Qual. Quant.* 45(2), 375–390 (2011)

Davison, R.M., Li, Y., Kam, C.S.P.: Web-based data collection in China. *J. Glob. Inf. Manag.* 14(3), 70–89 (2006)

Dwight, S.A., Feigelson, M.E.: A quantitative review of the effect of computerized testing on the measurement of social desirability. *Educ. Psychol. Meas.* 60(3), 340–360 (2000)

Fan, W., Yan, Z.: Factors affecting response rates of the web survey: a systematic review. *Comput. Hum. Behav.* 26(2), 132–139 (2010)

Fang, J., Wen, C., Pavur, R.: Participation willingness in web surveys: exploring effect of sponsoring corporation's and survey provider's reputation. *Cyberpsychol. Behav. Soc. Netw.* 15(4), 195–199 (2012)

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

French, B.F., Finch, W.H.: Multigroup confirmatory factor analysis: locating the invariant referent sets. *Struct. Equ. Model.* 15(1), 96–113 (2008)

Göritz, A.S.: Incentives in web studies: methodological issues and a review. *Int. J.*

Internet Sci. **1**(1), 58–70 (2006)

Ha, Y., Stoel, L.: Internet apparel shopping behaviors: the influence of general innovativeness. *Int. J. Retail Distrib. Manag.* **32**(8), 377–385 (2004)

Hair, J.F., Black, W.C., Anderson, R.E., Tatham, R.L.: *Multivariate Data Analysis*, 6th ed. Prentice-Hall, Upper Saddle River (1998)

Heerwegh, D., Loosveldt, G.: Personalizing E-mail contacts: its influence on web survey response rate and social desirability response bias. *Int. J. Public Opin. Res.* **19**(2), 258–268 (2007)

Higgins, L., Zheng, M., Liu, Y., Sun, C.: Attitudes to marriage and sexual behaviors: a survey of gender and culture differences in China and United Kingdom. *Sex Roles* **46**(3/4), 75–89 (2002)

Hurt, T., Joseph, K., Cook, C.: Scales for the measurement of innovativeness. *Hum. Commun. Res.* **4**(1), 58–65 (1977)

Keh, H.T., Sun, J.: The complexities of perceived risk in cross-cultural services marketing. *J. Int. Mark.* **16**(1), 120–146 (2008)

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

Lalwani, A., Chiu, C.: Cognitive load, need for closure, and socially desirable responding: cognitively constrained versus motivated response biases in cross-cultural consumer research. *Adv. Consum. Res.* **35**, 759–760 (2008)

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

Lu, J., Yao, J.E., Yu, C.S.: Personal innovativeness, social influences and adoption of wireless Internet services via mobile technology. *J. Strateg. Inf. Syst.* **14**(3), 245–268 (2005)

Mcknight, D.H., Choudhury, V., Kacmar, C.: Developing and validating trust measures for E-commerce: an integrative typology. *Inf. Syst. Res.* **13**(3), 334–359 (2002)

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

Miles, E.W., King, W.C.: Gender and administration mode effects when pencil-and-

paper personality tests are computerized. *Educ. Psychol. Meas.* **58**(1), 68–76 (1998)

Nunnally, J.C., Bernstein, I.H.: *Psychometric Theory*, 3rd edn). McGraw-Hill, New York (1994)

Paulhus, D.L.: Two-component models of social desirable responding. *J. Pers. Soc. Psychol.* **46**(3), 598–609 (1984)

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

Reise, S.P., Widaman, K.F., Pugh, R.H.: Confirmatory factor analysis and item response theory: two approaches for exploring measurement equivalence. *Psychol. Bull.* **114**(3), 552–566 (1993)

Rosen, P.A.: The effect of personal innovativeness in the domain of information technology on the acceptance and use of technology. A Working Paper (2004)

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., Schwartz, S.: Testing measurement invariance using multigroup CFA: differences between educational groups in human values measurement. *Qual. Quant.* **43**(4), 599–616 (2009)

Tellis, G., Chandrasekaran, D.: Extent and impact of response biases in cross-national survey research. *Int. J. Res. Mark.* **27**(4), 329–341 (2010)

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

Vandenberg, J., Lance, E.: A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ. Res. Methods* **3**(1), 4–70 (2002)

Vecchione, M., Alessandri, G., Barbaranelli, C.: Paper-and-pencil and web-based testing: the measurement invariance of the big five personality tests in applied settings. *Assessment* **19**(2), 243–246 (2012)

Westland, J.C.: Lower bounds on sample size in structural equation modeling. *Electron. Commer. Res. Appl.* **9**(6), 476–487 (2010)

Yetter, G., Capaccioli, K.: Differences in responses to web and paper surveys among school professionals. *Behav. Res. Methods* **42**(1), 266–272 (2010)

Zhang, N., Guo, X., Chen, G., Chau, P.: Impact of perceived fit on E-government user evaluation: a study with a Chinese cultural context. *J. Glob. Inf. Manag.* **17**(1), 49–69

(2009)