Computers in Human Behavior 30 (2014) 335–343

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

An assessment of equivalence between paper and social media surveys: The role of social desirability and satisficing

Jiaming Fang [a,*], Chao Wen [b,1], Victor Prybutok [c,2]

[a] Department of Management Science and E-business, University of Electronic Science and Technology of China, Chengdu, China
[b] School of Business, Eastern Illinois University, Charleston, USA
[c] Department of Information Technology and Decision Sciences, University of North Texas, Denton, USA

ARTICLE INFO

Article history:
Available online 17 October 2013

Keywords:
Social media
Social media surveys
Web survey
Measurement invariance
Social desirability
Satisficing

ABSTRACT

Investigation of the underlying mechanisms responsible for measurement variance has received little attention. The primary objective of this study is to examine whether paper and social media surveys produce convergent results and investigate the underlying psychological mechanisms for the potential measurement nonequivalence. Particularly, we explored the role of social desirability and satisficing on the measurement results. We collected data via five different survey modes, including paper survey, ad hoc Web survey, online forum (message boards)-based, SNS-based and microblog-based surveys. The findings show that socially desirable responding does not lead to inconsistent results. Rather we found that satisficing causes inconsistent results in paper versus online surveys. Sociability reduces the possibility of engaging in satisficing that results in inconsistent results between traditional Web surveys and social media-based Web surveys.

1. Introduction

An explosion of dedicated online communities and social network sites (SNSs) such as Facebook, Twitter, and LinkedIn is a hallmark of the last decade. As an interpersonal communications platform, the social media provides a venue where individuals meet their offline friends or others with similar interests and values. Perceived risk on the social media is relatively low for the frequent and pleasing contacts with other members and the affective commitment in relations (Casaló, Flavián, & Guinalíu, 2011).

The large number of users on these social media websites is gradually attracting the attention of academic and industry researchers who are starting to use social media survey platform. Individuals on social media websites are socially connected. Survey researchers can leverage the benefits of social media to increase a respondent's level of engagement, reduce recruiting costs, and acquire more credible insights. The participation rate of a social media survey mode is also expected to be higher than that of traditional Web surveys of which the non-response rate is believed to be too high.

Alternatively, the use of social media to collect data is restrained by coverage limitations for not including those who do not have access to the technology, and thus it is usually inappropriate to use the social media mode alone to conduct surveys. As a result researchers need to either limit the populations for which they use social media surveys or use multiple delivery methods. Practically, researchers usually adopt a mixed mode design in which data are collected by different survey modes to increase response rates, reduce survey costs, and alleviate the negative effects of a low coverage rate. However, the researchers need to know whether different survey modes can produce equivalent results when using a mixed mode design. If data collected from different survey modes produce different results for the same study, it is not appropriate to aggregate the results. Accordingly, when integrating data collected from social media-based surveys into traditional survey modes, researchers and practitioners must be aware and examine the measurement equivalence[3] of these survey modes and that the measured latent constructs have the same theoretical pattern under different survey modes (Miles & King, 1998). If the assumption of invariance is violated, it is not appropriate to combine the data gathered by disparate survey modes and conduct the analysis assuming a homogeneous data set.

* Corresponding author. Address: School of Management and Economics, University of Electronic Science and Technology of China, No. 4, Section 2, N. Jianshe Rd, Chengdu, Sichuan 610054, China.

E-mail addresses: jiamingfang@gmail.com (J. Fang), cwen@eiu.edu (C. Wen), victor.prybutok@unt.edu (V. Prybutok).

1 Address: School of Business, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920, USA.

2 Address: College of Business, University of North Texas, 1155 Union Circle #305249, Denton, TX 76203, USA.

3 In this paper, we will be using the terms 'measurement invariance' (MI) and 'measurement equivalence' (ME) interchangeably.

0747-5632/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chb.2013.09.019

There are several existing studies regarding the comparison of online and paper survey modes (e.g., Deutskens, de Ruyter, & Wetzels, 2006) and these studies have contributed to our understanding of MI between paper and online surveys. However, the online surveys in the existing MI research mainly refer to ad hoc Web surveys or online panel surveys. The social distance between participants and survey researchers in these surveys are usually large. Little empirical research explored whether paper and social media-based surveys produced comparable results and this study is proposed to fill this gap. With the exponential rising of a variety of social media, the sociability feature of the Internet is increasingly prominent and salient. Sociability refers to interpersonal affective connections embedded in social media which is derived from the online social interaction. Surveys based on social media can motivate individuals to actively participate in and potentially increase the respondents' candor. It is expected that the sociability embedded in social media surveys discourages respondents from engaging in satisficing behavior which is commonplace in traditional Web surveys. At the same time, the sociability may also compel respondents to exhibit more socially desirable responding (SDR). However, it is possible that social media surveys and traditional surveys such as paper surveys, ad hoc Web surveys and online panel surveys cannot provide comparable survey results. Therefore, it is critical to investigate whether these survey modes can produce convergent results as relevance to the emerging social media survey mode. An exhaustive literature review found no research devoted to investigating the potentiality of social media as a survey platform and examining the influence of sociability on the MI between social media surveys and paper surveys. This study contributes to the wide range of survey modes MI research by examining whether paper and social media surveys produce comparable results.

A second major contribution of this study relates to the limited insights of the existing MI research on potential psychological mechanism responsible for measurement variance across survey modes. Although a number of studies have examined the MI issues, most these studies simply reported the test results. Little effort has been devoted to investigating the potential underlying mechanisms responsible for the measurement variance. Especially lacking is the research on exploring psychological mechanisms for measurement nonequivalence. This study takes a step further and investigates the impacts of psychological factors such as SDR and satisficing on incomparable survey responses between paper surveys and social media-based surveys.

The following paper is organized as such. We begin with a brief introduction to social media-based surveys and describe the potential theoretical impacts of SDR and satisficing on MI between social media-based and paper surveys. Subsequently, we present the procedures for conducting MI tests, followed by data collection and data analysis to examine the impacts of SDR and satisficing on MI results. We conclude with a discussion of the results, research implications, and limitations.

2. Theoretical backgrounds

2.1. Social media-based surveys

Traditionally, the Internet is considered as a lean online communication medium, low in social presence, and imposes a range of risks associated with online transactions (Gutiérrez, Izquierdo, & Cabezudo, 2010; Wasko & Faraj, 2005). The emergence of social media is changing the lean nature of the online medium. Social media refers to the means of interactions among people in which they create, share, and exchange information and ideas in virtual online communities and networks.

The expanding participation on social media websites has encouraged researchers and consultants to identify ways they can take advantage of the application. One possible implication is to conduct social media-based surveys. Unlike traditional offline paper surveys and stand-alone surveys, social media surveys refer to online self-administered surveys conducted on the platforms of SNSs or using SNSs as survey URL distribution tools. For example, the message length of a microblog (such as Twitter) is usually constrained to 140 characters, and within that limit it is impossible to use Twitter as a survey platform. However, Twitter can serve as a tool for publishing an address hyperlinking to a survey website.

Although it is difficult if not impossible to acquire a representative sample of the general population by employing a social media survey design, social media samples are valuable for selecting market segments as well as when probability sampling is impossible to get. Social media surveys can expand the geographical scope and facilitate the identification of individuals with barriers to access, thus the use of social media can increase the sample size and representativeness (Baltar & Brunet, 2012). As an example, organizations can set up their own online communities or SNSs of customers and workers, and carry out surveys on these platforms to identify the needs of specific groups or recognize niche market opportunities.

More importantly, sociability embedded in social media is able to increase an individual's willingness to participate in surveys because of the lower level of perceived risks and relatively high participation motivation. Identity disclosure and relationship maintenance are two important functional building blocks of social media (Kietzmann, Hermkens, & McCarthy, & Silvestre, 2011). Previous research has shown that a survey researcher's identity and familiarity can significantly decrease perceived risks and affect respondents' perception about the salience of survey topics and their intention to participate (Fang, Wen, & Pavur, 2012). Sociability forms relational capital which is an important asset that benefits both the community and its members. Members are willing to assist other members, even strangers because of a strong sense of commitment and reciprocity (Xu, Ryan, Prybutok, & Wen, 2012).

2.2. Impact of SDR on MI

One existing explanation for the divergent results and measurement nonequivalence between paper and online surveys is that the Web platform offers relative anonymity to users. The anonymity associated with the Internet reduces the perceived risk. Researchers speculate that the Internet creates an impersonal social situation in which individuals feel more anonymous, more private, less inhibited, and less concerned about how they appear to others (Booth-Kewley, Larson, & Miyoshi, 2007). The reduced social context information on the Internet increases the outspokenness of online respondents and reduces tendencies to engage in SDR. SDR is the inclination to give answers that make the respondent appear good, or the propensity to make a good impression. Generally, respondents experience the desire to achieve greater social desirability in an environment where they are identified rather than anonymous (Paulhus, 1984). In traditional Website surveys such as ad hoc Website surveys and online forum-based surveys, respondents can choose to submit answers under conditions of anonymity. Thus, respondents may offer responses in these Website surveys different from those provided in offline surveys in which respondents always regard themselves as identifiable even if the promise of no identification cues included in questionnaires is given.

The existing research on SDR in different survey modes displays apparently conflicting results. Some research (e.g., Booth-Kewley et al., 2007) shows that SDR is more likely to occur in Web surveys than in paper surveys, while other research finds few or no SDR differences at all (e.g., King & Miles, 1995; Risko, Quilty, & Oakman, 2006). No research has been conducted to examine the SDR differences between paper and social media surveys or among online surveys. It is unclear whether the sociability embedded in social media surveys enables respondents to exhibit different extents of SDR, which can potentially result in measurement nonequivalence between social media surveys and other survey modes.

2.3. Impact of sociability on MI

Use of social media can provide the opportunity to develop and maintain social connectedness in the online environment (Grieve, Indian, Witteveen, Tolan, & Marrington, 2013). Sociability embedded in social media surveys can decrease the virtuality and increase the trust. Despite perceived anonymity exists in online environment, we speculate that potential respondents are more likely to be motivated to expend greater cognitive effort to fill in the questionnaires in social media surveys than in traditional Web surveys. It is the difference in cognitive efforts that leads to measurement variance across survey modes. A theoretical framework that can support our speculation is the satisficing theory (Krosnick, 1991).

According to the satisficing theory, respondents with low motivation are likely to engage in a suboptimal response strategy (satisficing) instead of using the optimal one (Kaminska, Mccutcheon, & Billiet, 2010). Satisficing is in turn likely to result in measurement nonequivalence across survey modes. 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 US engage in the satisficing behavior in surveys (Barge & Gehlbach, 2012).

The Internet provides few social ties, especially in ad hoc Web surveys. The lack of strong social ties between survey respondents and survey researchers can lead respondents to expend less energy in comprehensively understanding a survey questions, searching their memories for an optimal response, and selecting a carefully weighed response choice especially in the situation where effort is needed to finish a survey with little or no apparent reward from survey researchers. In addition, compared to an offline environment, the sense of disinhibition on the Web increases the likelihood of respondents' engaging in multitasking. Online survey respondents are more easily distracted by such things as listening to music, watching videos, chatting with friends, and playing online games while they are taking surveys. Thus, respondents are less likely to expend the required cognitive effort to answer questions carefully compared with the respondents in paper surveys. As a result, the respondents of traditional Web surveys are more likely to exhibit the satisficing behavior, thereby leading to measurement nonequivalence.

However, a participant's motivation in social media surveys is stronger than that in traditional Web survey modes. Motivation refers to the need for cognition (the willingness of respondents to expend time thinking) and the attitude toward surveys. Social media provides communities in which individuals meet their familiar offline friends or others with similar interests and values. Users can acquire a sense of community and involvement by frequent participations and contributions. In addition, conforming to the norm of reciprocity and maintaining social capital can also be used to explain the reasons why sociability associated with social media can encourage users to give assistance to the member who seeks help on the platform. Taken together, respondents in social media surveys are more easily motivated to expend a substantial amount of mental effort to fill out survey questionnaires.

Since no empirical research thus far addresses whether sociability exerts any impacts on measurement nonequivalence across different survey modes, the actual implications of sociability are unknown. The present research extends previous studies by assessing MI between paper and social media surveys, and investigating the potential underlying mechanisms by which the sociability exerts impacts on the MI.

3. Methodology

3.1. Measures

Two unidimensional innovativeness scales were selected for examination across the survey modes. One was the personal innovativeness in the domain of IT (PIIT), which is used to describe willingness to adopt a new information technology and is one of the most frequently used scales in information technology and information systems research. The other was Global Innovativeness (GI) scale, which assesses an individual's general predisposition to innovativeness, or willingness to try new things. Considering that PIIT belongs to the domain of a specific innovativeness scale, it is expected that there is a positive correlation between GI and PIIT.

The structure of the scale is depicted by Fig. 1. The GI scale consists of three measurement items ("I am generally open to accepting new ideas."; "I am willing to try new things."; "I feel that I am an innovative person.") borrowed from Clark and Goldsmith (2006) and the PIIT scale consists of four items ("If I heard about a new information technology, I would look for ways to experiment with it."; "In general, I am hesitant to try out new information technologies."; "Among my peers, I am usually the first to try out new information technologies. "; "I like to experiment with new information technologies.") borrowed from Agarwal and Parsad (1998).

The reasons why these two scales were chosen are based on the following considerations. First, the two scales are well-validated instruments with satisfactory psychometric properties and have simple structures. The simple structures enable respondents to easily understand the scales, thus helping to avoid the potential misunderstandings. In addition, short scales are less likely to induce boredom or fatigue in volunteer participants. Second, the scales involve in attitudes or opinions measurements. Most existing MI research focus on the scales measuring noncognitive skills such as personality, depression, and IQ. However, the social media such as Twitter, Facebook, and Google+ have become a major form of communication, and the expression of attitudes and opinions, for the general public. Individuals are willing to express their personal attitudes or opinions on these social media. Therefore, attitudes and opinions measurements are more appropriate and easier to be conducted on social media platforms. Third, prior MI research (e.g., Barbeite & Weiss, 2004) indicates that the self-assessment scales such as the computer self-efficacy scale, the computer anxiety scale and innovativeness scales are more likely to be affected by survey modes than the scales assessing an object's features (e.g. the computer playfulness scale, the perceived usefulness scale and the ease of use scale). The scales measuring an object's characteristics are more likely to obtain consistent results across survey modes. Thus, by using innovativeness scales we can obtain the relatively conservative results about the MI between paper and social media surveys. Taken together, the two innovativeness scales are appropriate and representative for the purpose of the study.

SDR is usually conceptualized as consisting of two components: impression management (IM) and self-deceptive enhancement (SDE) (Booth-Kewley et al., 2007). 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. In this study, the Paulhus Balanced Inventory of Desirable Responding (BIDR) was used to

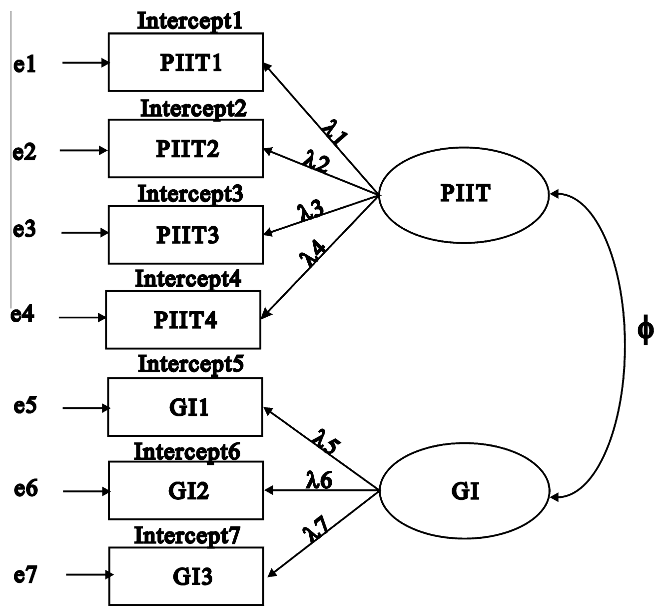

Fig. 1. The model for measurement invariance testing. PIIT represents personal innovativeness in the domain of IT; GI presents global innovativeness; e1–e7 represent measurement errors of the indicators; λ1–λ7 represent factor loadings relating the indicators to latent constructs of PIIT and GI; ϕ represents variance/covariance matrix of PIIT and GI.

assess the social desirability effects in each survey mode. It is composed of 40 items, the first 20 items for the SDE scale and the next 20 items for the IM scale. We included the SDR measurement in each survey mode.

There is no consensus among researchers regarding how to measure the level of satisficing. Previous research has normally measured satisficing using a single indicator such as middle response (Krosnick, Narayan, & Smith, 1996), extreme response (Belli, Herzog, & Van Hoewyk, 1999) and numbers of "don't know" responses (Kaminska et al., 2010). In this research, we assessed satisficing using two indicators. An "extreme + middle" metric (EM indicator) 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 larger magnitude of the EM metric indicates the higher possibility of satisficing. A second metric of inconsistent responses (INCR indicator) was used to gauge the extent of answer variation within a respondent. Prior research shows that satisficers do not select the exact same response for items but insufficiently adjusted their responses from one item to the next (Barge & Gehlbach, 2012). 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. A smaller value of the INCR metric represents the higher likelihood of satisficing.

3.2. Survey modes selection

We administered the surveys via paper survey, ad hoc Web survey, online forum (message boards)-based, SNS-based and microblog-based surveys. The disparate strengths of social ties are embedded in ad hoc Web surveys, forum-based, SNS-based, and microblog-based surveys.

An ad hoc Web survey is frequently done without plans for follow-up surveys. The connections between survey respondents and surveyors in ad hoc Web survey are temporary, distant, and fragile, because they are linked but not always for shared interests or activities, but for a utilitarian purpose. Thus, the sociability in ad hoc Web surveys is low. Forums are not usually implemented on authentic name profiles, and users always use avatars to represent themselves. In general, the social ties between individuals on forum are loose (Ellison, Steinfield, & Lampe, 2007).

Unlike ad hoc Web surveys and online forum surveys, prior research found that most SNSs primarily support preexisting social relations. For example, Facebook is mainly used to maintain existing offline relationships or solidify offline connections, as opposed to meeting complete strangers (Ellison et al., 2007). The relationships among individuals in an SNS are usually tightly-knit and emotionally close. Therefore, the social ties embedded in an SNS are much stronger than those in an online forum. Ties in microblogs are uni-directional. Relationships on microblogs are multiplex; users of microblogs can follow any one including close friends or totally unfamiliar individuals. The relationships are informal and without structure (Ellison et al., 2007). The sociability embedded in microblogs should be stronger than that in an online forum but weaker than that in an SNS.

3.3. Data collection

Students from a large public university in China were recruited to participate in five surveys. Respondents were encouraged to send invitation messages of doing survey to their connections in social media survey modes.

In the paper survey group, respondents completed the questionnaires in a quiet, large-group, classroom setting. We distributed 200 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.

Web-based surveys were hosted via a well-established online survey hosting site (http://www.sojump.com) in China. Sojump.com was founded in February 2004 and is similar to SurveyMonkey.com or Qualtrics.com in the U.S. More than 2.6 million surveys have been hosted on the site. E-mail addresses for the potential respondents were obtained in advance and the potential student respondents were informed that the survey URL would be sent using email invitations. 228 e-mails were sent and in two weeks we received 196 responses among which 137 responses were valid. The sample consisted of 59 male and 78 female respondents. Most of the respondents (81.2%) were between 18 and 28 years of age.

Forum-based Web surveys were conducted using a campus online forum at the same university. The forum was characterized by the real name registration, and each student has only one account. We created a new account and published a unique URL on this forum rather than hosting a survey questionnaire. The questionnaire was also placed on the Sojump.com Website. We received 177 responses to this anonymous survey and, after eliminating unusable data records, obtained a final sample of 129 participants. The sample consisted of 108 male and 21 female respondents. Most of the respondents (99.2%) were between 18 and 28 years of age.

We chose the Sina microblog (http://weibo.com) as a representative microblogging platform. Sina microblog (similar to Twitter) has over 600 million users and is one of the largest microblogging Websites in China. We included a unique URL address in a survey invitation message and potential respondents were linked to a questionnaire on the Sojump.com Website. Within a month, we received 166 responses of which 144 were valid. The sample consisted of 68 male and 76 female respondents. Most of the respondents (91.7%) were between 18 and 28 years of age.

We chose Renren (http://www.renren.com/) as the SNS survey platform. Renren (similar to Facebook) is one of the largest SNSs in China. Its users have exceeded 200 million. Since there is no online survey functional module on Renren as other SNS Websites, it is infeasible to use it as a hosting platform for the survey. As a result, in this study the SNS is also served as a survey URL distribution tool. We created the questionnaire on the Sojump.com Website and shared a Web link (URL) through Renren to get responses. We received 142 responses and 135 were valid. The sample consisted of 73 male and 62 female respondents. Most of the respondents (73.3%) were between 18 and 28 years of age. 28.2% of the respondents are undergraduates.

We adopted Westland's method (2010) of calculating the ratio of indicators to latent variables to verify the adequacy of the sample size to conduct MCFA. The results showed that the sample sizes in the five survey modes were acceptable to use for MCFA analysis.

To evaluate the demographic similarity among the five survey samples, we made comparisons on age, gender, and education level. A series of chi-square tests on age and education levels found no statistical difference; However, the composition of gender was significantly different among the five groups ($\chi^2(4) = 60.74$, $p < 0.01$).

3.4. Testing procedures

Consistent with prior research, we initially examined the equality of the observed variance–covariance matrices. If the matrices were invariant across the survey modes, additional analysis to test for MI was unnecessary. Otherwise, we further assess the equality of factor pattern matrices (configural invariance or weak factorial invariance) to examine whether the assumption of equality in covariance matrices was not satisfied. Next, we tested whether the factor loadings matrices were identical across the survey modes (metric invariance or strong factorial invariance). Then, the equality of the variance–covariance matrix among the underlying factors was further assessed if metric invariance was maintained. In this step, factor covariance was of the most interest with respect to the equality of theoretical structure (Raju, Laffitte, & Byrne, 2002). Finally, we evaluated the invariance of item intercepts across the survey modes (scalar invariance).

The chi-square goodness of fit and four other fit indices (CFI, TLI, SRMR, and RMSEA) were used to evaluate the goodness of fit of the general and nested models. These indices are sensitive to a lack of MI and were frequently used in past MI research (Chen, 2007). As rules of thumb, CFI and TLI values should exceed 0.90 to indicate a good fit, the RMSEA value should be less than 0.08. Although RMSEA is a commonly used fit measure, it does not behave well. In simulation studies, RMSEA over-rejects true models for small N (N is less than 250). By contrast, the SRMR is characterized as more sensitive to model misspecification than to sample size or violations of distributional assumptions (Hu & Bentler, 1999). Thus, SRMR is preferred in the current research. For a good model, the SRMR should be less than or equal to 0.09 (Iacobucci, 2010).

4. MI analysis and results

4.1. Descriptive data analysis

Exploratory factor analysis (EFA) was first conducted to determine if the number of factors for each scale measured in the current samples was consistent with those found in prior studies. Consistency in number of factors indicates a constant conceptual domain (Barbeite & Weiss, 2004). The results showed that only two factors could be extracted from each survey mode.

The Doornik-Hansen test (2008) for multivariate normality was conducted to confirm the assumption that the manifest variables should follow a multivariate normal distribution. The results showed that there were significant deviations from multivariate normality ($p < 0.01$) in the data sets of the five survey modes. Further, we performed multiple-sample multivariate tests on means, allowing heterogeneous covariance matrices across five survey modes. The results showed that *Wald* $\chi^2(28) = 67.33$ ($p < 0.01$), which indicated that the means of the observed items in the five survey modes significantly differed. Because of the assumption violation we used a robust ML estimator (MLR) instead of ML estimator.

A CFA model was specified for each sample from the five survey modes using a robust ML estimator with Yuan–Bentler corrections for non-normality data to test the consistency of factor loadings among items for each scale in different survey modes. The fit indices were satisfactory for each sample's data (MLR chi-square test statistics ranged between 18.90 and 32.77; CFI ranged between 0.94 and 0.98; TLI ranged from 0.90 to 0.96; SRMR ranged from 0.04 to 0.06). We evaluated the reliability of our constructs through composite reliability and average variance extracted (AVE). The results are provided in Table 1.

Table 1 shows that the composite reliability ranged between 0.82 and 0.92; thus, all values exceeded the recommended cutoff point of 0.70. The AVEs ranged from 0.54 to 0.77, so all constructs exceeded the 0.50 cutoff value. In addition, we compared the square root of AVE of each construct to the correlations with other constructs to evaluate the discriminant validity. The results showed that the square root of the AVE of each construct was greater than the correlations with other constructs. The results indicated that discriminant validity was met.

4.2. MI assessment

The MI tests were based on the analysis of mean and covariance structures, within the framework of CFA. We first evaluated equality of the covariance matrices across five survey modes by conducting an omnibus test of equality of variance–covariance matrices. If the matrices do not differ across groups, then MI is established, and further tests of other aspects of the ME are unnecessary. The results showed that $\chi^2_{YB}(112) = 186.99$ ($p < 0.01$), RMSEA = 0.07, SRMR = 0.18, which indicated that there were significant differences in the variance–covariance matrices across the five survey modes, so it was necessary to identify the source of nonequivalence.

The unequal variance–covariance matrices of observed items made determining whether the configural invariance held across survey modes necessary. Since a testable mean structure model requires fewer intercept and mean parameters in the model than there are means of observed variables (Bentler, 2005), we

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

Items	Paper modes			Online forum modes			Website modes			Microblog modes			SNS modes		
	C.R./AVE	*M*	*S.D.*	*C.R./AVE*	*M*	*S.D.*	*C.R./AVE*	*M*	*S.D.*	*C.R./AVE*	*M*	*S.D.*	*C.R./AVE*	*M*	*S.D.*
GI1	0.86/0.67	5.59	0.99	0.89/0.73	5.46	1.17	0.89/0.73	5.69	1.09	0.87/0.69	5.42	1.17	0.91/0.77	5.64	1.26
GI2		5.58	1.08		5.40	1.11		5.66	1.10		5.47	1.24		5.59	1.16
GI3		4.53	1.27		4.55	1.28		4.94	1.21		4.64	1.44		4.88	1.30
PIIT1	0.82/0.54	4.69	1.31	0.90/0.68	4.94	1.25	0.85/0.58	4.62	1.37	0.87/0.62	4.43	1.53	0.92/0.75	4.48	1.42
PIIT2		4.96	1.39		5.19	1.30		5.04	1.37		4.74	1.66		4.80	1.57
PIIT3		3.28	1.25		3.74	1.53		3.48	1.60		3.57	1.56		3.40	1.48
PIIT4		5.40	1.16		5.35	1.24		5.36	1.45		4.99	1.41		4.96	1.57

constrained the means of two latent variables in the five survey modes to zero to circumvent the identification problem. This model (Model 1) provided the baseline model value for comparison with subsequent tests for invariance. The results showed that $\chi^2_{YB}(65) = 128.35$, RMSEA = 0.08 , CFI = 0.96, TLI = 0.93, SRMR = 0.05. The fit indices indicated that we could not reject the null hypothesis that the equality of the number of factors and factor pattern matrices were held across the survey modes.

Because configural invariance had been established, we further evaluated metric invariance across survey modes (Model 2). Testing for the differences in factor loadings matrices resulted in a good model fit ($\chi^2_{YB}(85) = 156.68$, RMSEA = 0.08 , CFI = 0.95, TLI = 0.94, SRMR = 0.08). Since Model 2 was nested in Model 1, a Satorra–Bentler scaled Chi-square difference test was conducted to check whether the change of Chi-square was significant. The results showed that there was no difference between Model 1 and Model 2 ($\Delta\chi^2(20) = 28.06$, $p = 0.11$). Therefore, metric invariance was established. Accordingly, we proceeded in testing for the invariance of the covariance matrix among the factors.

In order to assess the invariance of the covariance matrix, we constrained the covariance to equality based on Model 2 while keeping the variances of factors freely estimated (Model 3), which led to $\chi^2_{YB}(89) = 168.87$, RMSEA = 0.08 , CFI = 0.95, TLI = 0.94, SRMR = 0.13. Comparison of Model 3 with Model 2 yielded a statistically significant difference ($\Delta\chi^2(4) = 12.84$, $p = 0.01$). The result showed that the theoretical structure of the latent variables was not the same across the survey modes. Data analysis revealed that the covariance between GI and PIIT was 0.23 in the paper survey mode, 0.61 in the Web survey mode, 0.54 in the online forum survey mode, 0.60 in the microblog survey mode, and 0.68 in the SNS survey mode. The covariance of the latent variables in the paper survey mode was much smaller than those in the other survey modes.

We found that the nonequivalence originated with the data set collected by the paper survey. Thus, we dropped this dataset, and then reanalyzed the MI across the remaining four survey modes. The results are shown in Table 2.

Table 2 shows that the four data sets from the four survey modes possessed equivalent theoretical structures. The variances and the relations among the latent factors were equal across the four survey modes. However, intercepts of the same items were nonequivalent across the survey modes which imply that the measurement items or the means of latent variables were varying across survey modes. Subsequent refinement in examination of the results via multiple attempts revealed that scalar invariance only existed between the data sets from the microblog and SNS survey modes.

4.3. Robustness checks

The demographic similarity test on the five survey samples showed that the proportion of males to females varied greatly. Previous studies indicated that women were more likely to have higher innovativeness than men (Ha & Stoel, 2004), and that males might be more affected by survey administration mode than females (Booth-Kewley et al., 2007). Therefore, it is necessary to confirm whether the MI test findings are caused by the dissimilar gender composition rather than the survey modes. We examined the impacts of gender on items within and across survey modes. A series of Wilcoxon rank-sum tests showed that gender did not exert consistently significant impacts on the observed items. Therefore, the differences of observed items' means are not caused by the differences in the gender proportions, and the measurement nonequivalence is attributed to using the disparate survey modes.

5. The roles of SDR versus satisficing on measurement results

5.1. SDR on measurement nonequivalence

Building upon prior research, we examined social desirability effects on measurement nonequivalence to reveal the underlying mechanisms of sociability on measurement nonequivalence. Since prior research shows that gender is potentially associated with the impression management and self-deceptive enhancement, thus we included gender as one of the independent variables. A multiple ANOVAs were conducted to test the main effects of survey modes, gender, and their interaction on the SDR behavior.

The results showed that there was no interaction effect between gender and survey mode on BIDR SDE, $F(4,688) = 0.31$, $p = 0.87$. The main effect of survey mode was also not significant, $F(4,688) = 0.80$, $p = 0.52$. A significant main effect of gender was found, with males scoring significantly higher on SDE than females, $F(1,688) = 13.0$, $p < 0.01$. Further, we integrated the four online survey modes into a combined group versus paper survey modes as another group. When comparing the combined online modes versus the paper mode the main effect of survey mode was insignificant, $F(1,694) = 1.78$, $p = 0.18$. For BIDR IM, the survey mode by gender interaction was not significant, $F(4,688) = 1.96$, $p = 0.10$. The main effect of survey mode was also not significant, $F(4,688) = 0.97$, $p = 0.42$. However, the gender main effect was significant with females consistently scoring higher on IM scores than males across the five survey modes, $F(1,688) = 16.81$, $p < 0.01$. When we integrated the four online survey modes into a group, paper survey modes as another group, the interaction effect of survey mode with gender was still insignificant, $F(1,694) = 0.05$, $p = 0.81$. The main effect of survey mode was also insignificant, $F(1,694) = 3.12$, $p = 0.08$.

5.2. Satisficing on measurement nonequivalence

Prior research shows that satisficers' response patterns are likely to exhibit bloated correlations between two variables (Barge & Gehlbach, 2012). In the present study, the correlation between GI and PIIT was much larger in Web surveys than in the paper survey. Thus, we further examined whether the levels of satisficing differed in different survey modes. Considering that gender might be associated with satisficing, gender was included as one of the independent variables.

Table 2
MI test of data sets from Web, forum, microblog, and SNS survey modes.

Model invariance	χ^2_{YB}	*df*	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	*p*
Variance–covariance of measurement items	138.76	84	0.96	0.96	0.07	0.14	–	–	<0.01
Configural invariance	100.97	52	0.96	0.94	0.08	0.05	–	–	–
Metric invariance	115.19	67	0.96	0.95	0.07	0.07	13.52	15	0.56
Covariance invariance	116.28	70	0.96	0.96	0.07	0.07	0.83	3	0.84
Variance–covariance of latent variables	126.71	76	0.96	0.96	0.07	0.12	10.34	6	0.11
Intercept invariance	165.13	97	0.95	0.95	0.07	0.15	38.74	21	0.01

Note: χ^2_{YB} = Chi-square with Yuan–Bentler corrections for non-normality data; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

We first assessed whether there was a satisficing difference between the paper survey and online surveys. The results showed that in the EM indicator and the INCR indicator the interactions of survey modes and gender were insignificant. The main effect of gender was also insignificant in the EM indicator, but was significant in the INCR indicator with females exhibiting more inconsistent responding, $F(1,659) = 7.03$, $p < 0.01$. However, the main effect of survey mode was both significant in the EM and INCR indicators (EM: $F(1,677) = 7.90$, $p < 0.01$; INCR: $F(1,659) = 21.39$, $p < 0.01$). The respondents completing the online surveys manifested significantly higher satisficing than those in the paper survey group. Accordingly, the satisficing in Web surveys results in a bolstered covariance between the two scales, which essentially leads to measurement variance between paper and online surveys.

Then, we inspected whether there were satisficing differences across online surveys. The Website and the online forum surveys were combined into a combined group (traditional online survey group), and the SNS and microblog surveys were combined into another group (social media group). The results showed that, in the EM indicator, both the interaction and main effects were insignificant. In the INCR metric, the interaction and gender main effects were insignificant, but the main effect of survey mode was significant, $F(1,520) = 4.36$, $p = 0.04$. The respondents in the traditional survey group exhibited higher satisficing than those in the social media survey group.

Finally, we examined whether there was a satisficing difference between paper and social media surveys. The results showed that both in the EM indicator and the INCR indicator the interactions and gender main effects were insignificant, but the main effects of survey mode were significant (EM: $F(1,411) = 7.32$, $p < 0.01$; INCR: $F(1,399) = 11.19$, $p < 0.01$). The respondents in social media surveys demonstrated relatively higher satisficing than those in the paper survey.

6. Discussion

Mixed-mode surveys are increasingly used by both academics and practitioners to obtain sufficient responses, access a broader base of respondents, acquire representative samples, and reduce the survey costs. Nevertheless, combining data collected from different modes is potentially problematic if respondents answer the surveys differently on different modes.

The chief objective of this research is to assess the MI between paper and social media surveys and explore how the sociability embedded in social media surveys affects the measurement results. The research results show that survey modes exert an impact on measurement differences. However, there is essentially no difference in SDR among survey administration modes. Previous research contends that social context information in online surveys is weak, which increases the respondents' perceived anonymity and thus produces relatively self-centered and unregulated behavior (Deutskens et al., 2006). Therefore, respondents may be less prone to give socially desirable answers in Web surveys. The differences in SDR lead to measurement nonequivalence between paper surveys and online surveys (Booth-Kewley et al., 2007). However, those research findings are obviously inconsistent as a justification for the nonequivalence in this study. No effects of SDR on the measurement nonequivalence due to administration mode were found on either of the BIDR scale (IM and SDE).

Our results showed that compared to the paper surveys, respondents were more inclined to satisfice on Web surveys. The lean medium Internet provides few social ties in traditional Web surveys, while paper surveys have more social ties for the physical contact between respondents and surveyors. Moreover, online surveys offer respondents a greater chance for multitasking than in paper surveys. Online survey respondents are more easily distracted by such things as listening to music, chatting with friends, and playing games whilst they are taking surveys. Respondents are more likely to squeeze the survey in between two other commitments (Barge & Gehlbach, 2012). As a result of the disruption for attention, fewer respondents to Web surveys will dedicate the time and attention to the survey than researchers would expect.

Our results further provided some evidence that sociability embedded in social media surveys could exert an influence on the survey results. Sociability to some extent reduces the possibility of engaging in satisficing behavior in Web surveys, which in turn results in measurement nonequivalence between traditional Web surveys and social media surveys. However, our results also indicated that sociability embedded in social media surveys could not totally eliminate the satisficing behavior in the online environment. The respondents in social media surveys showed significantly higher satisficing behavior than those in the paper survey. One possible explanation for the relatively weak impact of sociability may be due to the anonymous nature of the Web surveys conducted in the current study. Although respondents are more likely to be motivated to devote adequate cognitive effort to answering questions carefully in social media surveys, the anonymous survey environment weaken the motivation. The top two reasons why people join social networking sites are to showcase identities and to connect with others (Kietzmann et al., 2011). Anonymity destroys the dynamics of aggregation on social media platforms. Research by Burnett and Illingworth (2008) provided another support to our claim. Their research showed that in an anonymous environment, compared to an identified environment, individuals are unwilling to devote time to engaging in knowledge sharing and are reluctant to give suggested improvements and helpful comments. Our results indicated that when social media surveys were conducted in an identified environment, the likelihood of engaging in satisficing might decrease.

The results also provided some indication that females were less likely to engage in satisficing in both online and offline surveys. One possible explanation is that women feel a stronger sense of responsibility to help others. Another potential explanation is that women exhibit a higher level of conformity in offline contexts and in an anonymous computer-mediated communication context (Lee, 2006). We also suspect that the finding may be specially related to Chinese culture and Confucian dynamism (Shin, Ishman, & Sanders, 2007). Women are considered inferior to men per Confucian tradition and are lower than men in the social hierarchy; thus, women are more reluctant to engage in risky or otherwise questionable behavior in comparison to men (Hovav & D'Arcy, 2012). People in China value harmony and moderation, and try to conform to the expectations of others. Women are more likely to commit to the social norms than men and exhibit greater conformance than men. Therefore, females are more likely than males to expend a great deal of time and mental effort to complete surveys.

The findings of this study are significant because of their implications for both researchers and practitioners who intend to use social media-based surveys as a portion of a mixed-mode survey design. Although the results of several existing MI studies between online surveys and paper surveys are insightful, online surveys in those studies usually refer to traditional online surveys that do not rely on social media to collect data. The social distance between survey researchers and respondents are usually large in traditional Web surveys. Sociability embedded in the emerging social media-based surveys reduces the virtuality and increases trust between survey researchers and respondents. Therefore, it is invaluable to investigate whether and how sociability impacts the measurement results. There are two major contributions of this research. First, we took the initial step to address the MI issue between paper surveys and social media-based surveys. Second, our study investigated both SDR and satisficing as mechanisms responsible for the measurement variance across survey modes. Prior research suggests that it is the SDR that affects measurement nonequivalence between paper surveys and online surveys because the online environment renders respondents a strong feeling of disinhibition. However, the speculation does not consider the influences of sociability and is unable to explain the reason why response differences exist among online surveys. Drawing on the theory of satisficing, we included satisficing as a mechanism by which the inconsistent results occur in mixed-mode surveys. The results identified satisficing as the main reason leading to the inconsistent answers between traditional online surveys and social media surveys, and between paper surveys and Web surveys.

Beyond the theoretical contribution of the study, there are some managerial implications of the findings. This study revealed similar SDR differences between online surveys and paper surveys. However, it is implausible to eliminate the systematic tendency of some respondents to present themselves in a favorable manner (Baumgartner & Steenkamp, 2005). Controlling SDR as a response style, researchers can include an SDR scale as part of the total survey instrument and add respondents' scores on the SDR instrument as a control variable in analyses to control the SDR contamination in respondents' responses.

More importantly, the research results suggest that when collecting data via a mixed-mode survey approach, concern should be raised on the satisficing behavior in both online surveys and offline surveys. Satisficing might improve measurement reliability and validity but it can also produce artificially high associations between the latent variables. The multivariate analysis methods such as the structural equation modeling and factor analysis that are commonly employed in the behavioral research are largely based on correlation matrices or covariance matrices. Therefore, examining satisficing should be a routine in the data preprocessing stage. Interactive intervention can be used as an effective strategy to curtail respondent satisficing behaviors (e.g., Holland & Christian, 2009). Intervention design in online surveys such as popping up a window with a warning message to discourage the satisficing behavior can get very promising results. However, in the traditional online surveys the interactive feedback about respondents' behaviors may produce more reports of SDR answers compared to the no-intervention condition.

7. Limitations

Several limitations of this work should be noted. First, in the current research we constrained social media surveys to microblog and SNS surveys, and the social media was only employed as a survey URL delivery platform. Thus, our results may not generalize to other forms of social media surveys and the situations in which social media is used as platforms for placing the questionnaire. Second, in the current research the term Web survey refers to ad hoc Web surveys or stand-alone Web surveys. We did not consider the situation in which online panels were used in surveys. In ad hoc Web surveys, since follow-up surveys are not generally conducted on the same sample, the relations between participants and survey researchers are temporary and weak. In such an environment, survey researchers usually have to offer incentives to increase the response rates because of the weak social ties and the high perceived risk in ad hoc Web surveys. Compared with online panel surveys, since panelists who are invited to a survey sign up with the panel beforehand to take part in several studies, a certain degree of trust has been established between respondents and survey researchers. Thus, the findings of this research may not be proper to be applied to this situation. Finally, because our samples mainly consisted of university students, the results may not be generalized to other surroundings.

8. Conclusions

MI is an important precondition to mix the data from the different survey modes. There is no research conducted to address the equivalence of traditional paper surveys and emerging social media surveys or between traditional Web surveys and social media surveys. As an initial effort to fill this gap, this study explores the equivalence between traditional surveys and social media surveys and inspects two potential underlying mechanisms for the measurement nonequivalence. We establish that it is the satisficing rather than social desirability responding that provokes respondents to provide dissimilar responses between paper and online surveys. The results also show that sociability embedded in social media surveys reduces the possibility of engaging in the satisficing behavior in Web surveys that can result in measurement variance between traditional Web surveys and social media-based Web surveys.

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.

References

Agarwal, R., & Parsad, J. (1998). A conceptual and operational of personal innovativeness in the domain of information technology. *Information Systems Research, 9*(2), 204–215.
Baltar, F., & Brunet, I. (2012). Social research 2.0: Virtual snowball sampling method using Facebook. *Internet Research, 22*(1), 57–74.
Barbeite, F., & Weiss, E. (2004). Computer self-efficacy and anxiety scales for an Internet sample: Testing measurement equivalence of existing measures and development of new scales. *Computers in Human Behavior, 20*(1), 1–15.
Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education, 53*(2), 182–200.
Baumgartner, H., & Steenkamp, J.-B. (2005). Response biases in marketing research. In R. Grover & M. Vriens (Eds.), *The handbook of market research: Do's and don'ts* (pp. 204–237). Thousand Oaks, CA: Sage Publishing.
Belli, R. F., Herzog, R. A., & Van Hoewyk, J. (1999). Scale simplification of expectations for survival: Cognitive ability and the quality of survey responses. *Cognitive Technology, 4*(2), 29–38.
Bentler, P. M. (2005). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.
Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaire. *Computers in Human Behavior, 23*(1), 463–477.
Burnett, S., & Illingworth, L. (2008). Anonymous knowledge sharing in a virtual environment: A preliminary investigation. *Knowledge and Process Management, 15*(1), 1–11.
Casaló, L. V., Flavián, C., & Guinalíu, M. (2011). Antecedents and consequences of consumer participation in on-line communities: The case of the travel sector. *International Journal of Electronic Commerce, 15*(2), 137–167.
Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504.
Clark, R. A., & Goldsmith, R. E. (2006). Global innovativeness and consumer susceptibility to interpersonal influence. *Journal of Marketing Theory and Practice, 14*(4), 275–285.
Deutskens, E., de Ruyter, Ko., & Wetzels, M. (2006). An assessment of equivalence between online and mail surveys in service research. *Journal of Service Research, 8*(4), 346–355.
Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics, 70*(S1), 927–939.
Ellison, N., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends": Exploring the relationship between college students' use of online social networks and social capital. *Journal of Computer-Mediated Communication, 12*(3), 1143–1168.
Fang, J., Wen, C., & Pavur, R. (2012). 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.
Grieve, R., Indian, M., Witteveen, K., Tolan, G. A., & Marrington, J. (2013). Face-to-face or Facebook: Can social connectedness be derived online? *Computers in Human Behavior, 29*(3), 604–609.

Gutiérrez, S. M., Izquierdo, C., & Cabezudo, R. (2010). Product and channel-related risk and involvement in online contexts. *Electronic Commerce Research and Applications, 9*(3), 263–273.

Ha, Y., & Stoel, L. (2004). Internet apparel shopping behaviors: The influence of general innovativeness. *International Journal of Retail & Distribution Management, 32*(8), 377–385.

Holland, J. L., & Christian, L. M. (2009). The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. *Social Science Computer Review, 27*(2), 196–212.

Hovav, A., & D'Arcy, J. (2012). Applying an extended model of deterrence across cultures: An investigation of information systems misuse in the US and South Korea. *Information and Management, 49*(2), 99–110.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*(1), 90–98.

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

Kietzmann, J., Hermkens, K., McCarthy, I., & Silvestre, B. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons, 54*(3), 241–251.

King, W., & 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.

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., & Smith, W. (1996). Satisficing in surveys: Initial evidence. In M. T. Braverman & J. K. Slater (Eds.), *Advances in survey research* (pp. 29–44). San Francisco, CA: Jossey-Bass.

Lee, E.-J. (2006). When and how does depersonalization increase conformity to group norms in computer-mediated communication? *Communication Research, 33*(6), 423–447.

Miles, E. W., & 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.

Paulhus, D. L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598–609.

Raju, S., Laffitte, J., & 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.

Risko, E. F., Quilty, L. C., & Oakman, J. M. (2006). Socially desirable responding on the web: Investigating the candor hypothesis. *Journal of Personality Assessment, 87*(3), 269–276.

Shin, S. K., Ishman, M., & Sanders, G. L. (2007). An empirical investigation of socio-cultural factors of information sharing in China. *Information & Management, 44*(2), 165–174.

Wasko, M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly, 29*(1), 35–57.

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

Xu, C., Ryan, S., Prybutok, V., & Wen, C. (2012). It is not for fun: An examination of social network site usage. *Information & Management, 49*(5), 210–217.