

Association for Information Systems AIS Electronic Library (AISeL)

Eleventh Wuhan International Conference on e-Business

Wuhan International Conference on e-Business

5-26-2012

Measurement Equivalence of Web Surveys Based on Social Media

Wenlong Zhu

School of Management and Economics, University of Electronic Science and Technology of China, China, zhuwenlong1985@163.com

Peiji Shao

School of Management and Economics, University of Electronic Science and Technology of China, China, shaopj@uestc.edu.cn

Jiaming Fang

School of Management and Economics, University of Electronic Science and Technology of China, China

Follow this and additional works at: <http://aisel.aisnet.org/whiceb2011>

Recommended Citation

Zhu, Wenlong; Shao, Peiji; and Fang, Jiaming, "Measurement Equivalence of Web Surveys Based on Social Media" (2012). *Eleventh Wuhan International Conference on e-Business*. 86.
<http://aisel.aisnet.org/whiceb2011/86>

This material is brought to you by the Wuhan International Conference on e-Business at AIS Electronic Library (AISeL). It has been accepted for inclusion in Eleventh Wuhan International Conference on e-Business by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Measurement Equivalence of Web Surveys Based on Social Media

Zhu Wenlong¹, Shao Peiji¹, Fang Jiaming^{1*}

¹School of Management and Economics, University of Electronic Science and Technology of China, China

Abstract: Web surveys have become a very important way of empirical research. The rise of social media provides a new platform for web surveys. Investigators can collect data based on different social media platforms to get more respondents of higher external validity at a lower cost. However, there is no domestic expert and scholar to conduct related researches on web surveys based on social media. This paper studies the measurement equivalence of surveys among BBS, micro blog and SNS using CFA. The result shows that three surveys have the same basic structure, factor loading matrix and factor covariance matrix; moreover, the intercepts are equivalent across micro blog and SNS.

Keywords: web surveys, measurement equivalence, social media, Confirmatory Factor Analysis (CFA)

1. INTRODUCTION

Social media is a site and technology that allows people to write, share, appraise, discuss and communicate with each other ^[1], which has recently become the main media by which people expand their interpersonal network online. Currently, social media includes mainly social networking sites (SNS), micro blog, blog, podcasts, wikis, BBS, etc. In June 2011, the time that Internet users visited SNS exceeded the portal sites for the first time ^[2], which indicated that social media as a new information service model had been accepted by Internet users.

The rise and rapid growth of social media creates good conditions for web surveys. Nevertheless, because of there are a variety of social media forms, when conducting web surveys with different social media platforms, we need to focus on an important issue is whether surveys based on different social media platforms meet measurement equivalence. In other words, whether the same tools on different platforms have the same explicatory ability and measured effect on respondents. When comparing the difference among the groups, if we don't examine measurement equivalence, then any further data analysis could yield misleading results.

This study selects Global Innovativeness Scale and Personal Innovativeness of Information Technology Scale, and examines measurement equivalence of surveys based on BBS, micro blog and SNS using Confirmatory Factor Analysis (CFA). This paper tests the level of equivalence among three surveys, and determines survey modes with higher level of equivalence on this basis.

By empirical analysis, this article can provide academic support for deeply researches of subsequent scholars in this field, and also offer appropriate reference for innovation and exploration of the other direction. According to the results, investigators can integrate the data of three surveys together taking a certain way for more samples.

2. LITERATURE REVIEW

According to Drasgow and Kanfer ^[3], a test or a subscale is said to have measurement equivalence across groups or populations if persons with identical scores on the latent construct have the same expected raw score

* Supported by the National Natural Science Foundation of China (No. 71101018)

** Corresponding author. Email: shaopj@uestc.edu.cn(Shao Peiji), zhuwenlong1985@163.com(Zhu Wenlong)

or true score at the item level, the subscale total score level, or both.

So far, there have been many foreign experts and scholars to study and explore the issue of measurement equivalence. However, there is considerable inconsistency in foreign countries for measurement equivalence. Kreuter ^[4], Stephenson ^[5] and others thought two surveys had measurement equivalence. Fouladi ^[6], Bushman ^[7] and others indicated two surveys didn't have measurement equivalence. Watchravesringkan ^[8], Nien ^[9] and others believed these two models existed partial measurement equivalence.

Currently, the main foreign research of social media includes user research, basic technology, application research, privacy, trust and risk research, and each main research has some subfields. For example, Greene ^[10] studied the characteristics of information behaviors of users on Face book. Brandtzaeg ^[11] studied a variety of motives of social networking users. Clements ^[12] constructed annotated socially map using random walk model. Hoffman ^[13] devoted to the marketing research in social media. Dowdell ^[14] recognized and avoided certain risk behaviors using technical means.

Domestically few experts and scholars devoted to researches on measurement equivalence relative to foreign researches. Cai ^[15] examined measurement invariance between pen-and-pencil survey and web survey in the context of Chinese culture. Bai ^[16] tested out gender with measurement equivalence using social desirability scale. Fang ^[17] examined equivalence of surveys among pen-and-pencil, web and BBS. The researches focused on social media are mainly qualitative and theoretical, and empirical researches are mostly shallow, lacking of further studies.

Domestic studies are mostly concentrated on traditional survey modes. If researchers need to use social media as one of platforms to collect data, it is exceptionally important that we detect measurement equivalence of surveys among different social media platforms. Therefore, it is necessary to study this issue.

3. THEORETICAL BASIS

3.1 Research methods

Currently, there are two popular methods for establishing measurement equivalence: Confirmatory Factor Analysis (CFA) and item response theory (IRT) ^[18]. Given characteristics of research and operational adaptability, this study chooses CFA to conduct empirical research.

3.2 Testing procedures

According to the opinions of Byrne ^[19], Vandenberg and Lance ^[20], testing procedures of this article are as follows:

- Equivalence test of the observed variance-covariance matrices Σ , which is $H_{\Sigma} : \Sigma^{(1)} = \dots = \Sigma^{(3)}$.
- Equivalence test of the number of factors and factor pattern matrices, which is the construction of the baseline model.
- Equivalence test of the factor loading matrices Λ , which is $H_{\Lambda} : \Lambda^{(1)} = \dots = \Lambda^{(3)}$.
- Equivalence test of the factor variance-covariance matrices Φ , which is $H_{\Lambda, \Phi} : \Phi^{(1)} = \dots = \Phi^{(3)}$.
- Equivalence test of the intercepts τ , which is $H_{\Lambda, \Phi, \tau} : \tau^{(1)} = \dots = \tau^{(3)}$.
- Equivalence test of the variance-covariance matrixes among measurement errors Θ , which is $H_{\Lambda, \Phi, \tau, \Theta} : \Theta^{(1)} = \dots = \Theta^{(3)}$.

In addition to H_{Σ} , each step is the basis of the next step. When this step is met, other equivalence tests may be performed; once a certain step is not met, subsequent equivalence tests are considered unjustified.

4. DATA

4.1 Scales

This study selects two correlated scales to conduct equivalence test. One is Global Innovativeness (GI) ^[21] Scale, and the other is Personal Innovativeness of Information Technology (PIIT) ^[22] Scale, which is designed based on GI in the background of Internet.

The main reasons why we choose the two scales are as follows: firstly, two scales have been widely used in many fields, and they are both valid and reliable; secondly, the two scales are correlated, which is prone to examine the equivalence of the factor variance-covariance matrix; thirdly, measured contents are closely related to respondents, and the two scales are simple and easy to understand.

The two scales use 7-point Likert-type scale ranging from 1(strongly disagree) to 7(strongly agree). GI Scale includes 3 items and PIIT Scale has 4 items. The conceptual model of factors and measurement indicators is shown in Figure 1.

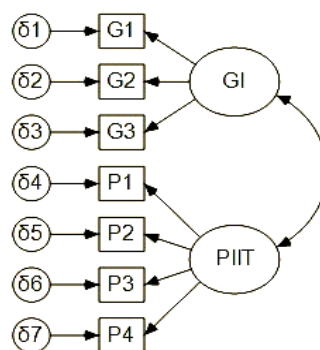


Figure 1. Conceptual model of factors and measurement indicators

4.2 Data collection

Given the feasibility and practical operability of this study, we choose BBS, micro blog and SNS as analytic platforms.

Based on the survey perspective, social media can be used for two purposes. It can be as URL platform and questionnaire platform. The former is mainly used to recruit subjects, the latter aims to place the questionnaire. When BBS as a questionnaire platform, we cannot delete questionnaires submitted by respondents and are unable to prevent the back viewing the previous results, so this study uses BBS as a recruit platform in order to ensure data quality. There are some restrictions in micro blog system, and therefore it is only used as a recruit platform. The status of the surveys and statistics is shown in Table 1.

Table 1. Status of the surveys and statistics

Platform	Mode	Obtained number	Effective number	Effective rate (%)	Proportion of males (%)	Proportion of 18-28 years old (%)	Survey status
BBS	Recruit	177	129	72.9	83.7	100.0	BBS of western university
Micro blog	Recruit	166	144	86.7	47.2	91.7	Sina micro blog
SNS	Recruit and distribute	142	135	95.1	54.1	73.3	Select ww.renren.com as recruit platform and QQ as place platform

In SNS surveys, recruitment survey collects 83 questionnaires and effective number is 78; distribution survey collects 59 questionnaires and effective number is 57. Using Box's test of the observed covariance matrix ^[23], we have $F=1.35$ ($p=0.10>0.05$), showing two types of surveys in SNS are consistent. So we can mix two types of data for further analysis.

To ensure the statistical similarity among three surveys, we conduct K-S test on gender and age to examine whether two nominal variables have an impact on the results. The results show that the means of seven items in different age groups and genders are both no significant difference.

4.3 Tool and fit indexes

We choose Lisrel 8.7 software as analytic tool, and use fixed loading method to establish the baseline model. We select χ^2 , NNFI, CFI, RMSEA to evaluate goodness-of-fit of subsequent models^[24].

5. ANALYSIS

In this study, we set factor loading of G1 item and P1 item to 1.0. In accordance with test procedures, the results are shown in Table 2. M' is the model with the same observed variance-covariance matrices; M0 is the baseline model; M1 is the model with the same factor loading matrices; M2 is the model with the same factor variance-covariance matrices; M3 is the model with the same intercepts.

Table 2. Results of equivalence test in three surveys

Model	Test	χ^2	df	NNFI	CFI	RMSEA	$\Delta\chi^2$	Δdf	P
M'	H_{Σ}	118.57	56	0.96	0.97	0.09	-	-	<0.001
M0	Model equivalence	87.31	39	0.96	0.97	0.09	-	-	-
M1	H_{Λ}	96.31	49	0.97	0.97	0.08	9.00	10	0.53
M2	$H_{\Lambda,\Phi}$	102.05	55	0.97	0.97	0.08	5.74	6	0.45
M3	$H_{\Lambda,\Phi,\tau}$	129.39	69	0.97	0.97	0.07	27.34	14	0.02

In the test of H_{Σ} , $p < 0.05$, indicating three surveys are not equivalent. So we should explore that three surveys meet the level of equivalence in a later step.

In the further testing procedures, the factor variance-covariance matrices of three surveys are equivalent, but the intercepts don't meet measurement equivalence. The result shows the starting points of the same scale in different platforms are unequal, and the estimated factor scores are biased.

According to the suggestions of Vandenberg and Lance^[20], on the basis of M2 we conduct a new testing, finding the intercepts in micro blog and SNS are equivalent. The detailed results are shown in Table 3. M4 is the model with the same variance-covariance matrices of measurement errors.

Table 3. Results of equivalence test in micro blog and SNS

Model	Test	χ^2	df	NNFI	CFI	RMSEA	$\Delta\chi^2$	Δdf	P
M'	H_{Σ}	63.90	28	0.96	0.97	0.09	-	-	<0.001
M0	Model equivalence	56.53	26	0.96	0.98	0.09	-	-	-
M1	H_{Λ}	60.07	31	0.97	0.98	0.08	3.54	5	0.62
M2	$H_{\Lambda,\Phi}$	61.89	34	0.97	0.98	0.07	1.82	3	0.61
M3	$H_{\Lambda,\Phi,\tau}$	67.93	41	0.98	0.98	0.07	6.04	7	0.54
M4	$H_{\Lambda,\Phi,\tau,\Theta}$	114.99	48	0.95	0.95	0.10	47.06	7	<0.001

We can learn that the intercepts of two surveys based on micro blog and SNS are equivalent ($p = 0.54 > 0.05$). There is no systematic bias when we implement surveys using the two platforms. However, the surveys between micro blog and SNS don't have the same variance-covariance matrices of measurement errors ($p < 0.001$).

6. RESULTS

After this empirical research, we can obtain two results.

- The factor variance-covariance matrices of three surveys are equivalent, which implies the baseline models are equivalent, $\Lambda^{(1)} = \dots = \Lambda^{(3)}$, $\Phi^{(1)} = \dots = \Phi^{(3)}$.

- The intercepts of surveys based on micro blog and SNS are equivalent, which implies the baseline models are equivalent, $\Lambda^{(1)} = \dots = \Lambda^{(3)}$, $\Phi^{(1)} = \dots = \Phi^{(3)}$, $\tau^{(1)} = \dots = \tau^{(3)}$.

According to the Chan and Schmitt^[25], when the factor loading matrices are invariance across groups, the measurement equivalence is established. In other words, differences of three platforms will not have a substantive impact on measured scales, and we could integrate the data of three surveys together.

Social media is a diversified platform, which represents the circle with interpersonal networks, showing the “social” features. In the social media with certain relationships, subjects will increase their trusts when they are aware of investigator identity, and will be harder to answer the questions.

The BBS, micro blog and SNS represent three different patterns of interpersonal relationships. The BBS this paper selected is a real-name system. The familiarity among individuals is weaker in this circle, so BBS belongs to the weak interpersonal social media. The interpersonal relationships in SNS are nearly the same as the realistic networks. The www.renren.com and QQ platform this article selected are both stressing the communication of “friend circle”, so SNS belongs to the strong interpersonal social media. The Sina micro blog this study selected is not only an extension of realistic networks, but also a new contact with the same information needs, so its relationships are between BBS and SNS. In this paper, we call it the mixed interpersonal social media.

The social network theory indicates that the relationships and communications among individuals play a more important role relative to individual characteristics^[26]. The results of this study are also good interpretation of this theory. The micro blog and SNS rose in 2008, and they were products of Web 2.0 technology, reflecting the strong social characteristics. Moreover, the information dissemination and interpersonal strength between them are more similar, thus they meet a higher level of measurement equivalence.

7. LIMITATIONS AND PROSPECTS

This paper is full of some significance, but this research still has many limitations. The issue of ordinal data and multivariate normality will bring estimated bias; the differences of sample size in each group will affect the accuracy of results; this article uses just BBS, micro blog and SNS as a representative of the entire social media, which will cause a certain degree of unreasonableness.

This paper studies the issue of measurement equivalence. Is the quality of data collected through the three modes the same? Are there significant differences? This will be our next work.

REFERENCES

- [1] Chen Y B, Fay S, Wang Q. (2011).The role of marketing in social media: how online consumer reviews evolve. *Journal of Interactive Marketing*, 25(2): 85-94
- [2] Kleiner Perkins Caufield & Byers (KPCB). (2011). *Internet Trends*. San Francisco: Web 2.0 Summit
- [3] Drasgow F, Kanfer R. (1985).Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70(4): 662-680
- [4] Kreuter F, Presser S, Tourangeau R. (2008).Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5): 847-865
- [5] Stephenson L B, Crête J. (2011).Studying political behavior: a comparison of Internet and telephone surveys. *International Journal of Public Opinion Research*, 23(1): 24-55

- [6] Fouladi R T, McCarthy C J, Moller N P. (2002).Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment*, 9(2): 204-215
- [7] Bushman T, Johnson J A, Goldberg L R. (2005).Implementing a five-factor personality inventory for use on the internet. *European Journal of Psychological Assessment*, 21(2): 115-127
- [8] Watchravesringkan K T, Yan R N, Yurchisin J. (2008).Cross-cultural invariance of consumers' price perception measures: eastern Asian perspective. *International Journal of Retail and Distribution Management*, 36(10): 759-779
- [9] Nien C L, Duda J L. (2008).Antecedents and consequences of approach and avoidance achievement goals: a test of gender invariance. *Psychology of Sport and Exercise*, 9(3): 352-372
- [10] Greene J A, Choudhry N K, Kilabuk E, et al. (2011).Online social network by patients with diabetes: a qualitative evaluation of communication Face book. *Journal of General Internal Medicine*, 26(3): 287-292
- [11] Brandtzaeg P B, Heim J. (2009).Why people use social networking sites. *Online Communities and Social Computing*, 5621: 143-152
- [12] Clements M, de Vries A P, Reinders M J T. (2010).The influence of personalization on tag query length in social media search. *Information Processing & Management*, 46(4): 403-412
- [13] Hoffman D L, Fodor M. (2010).Can you measure the ROI of your social media marketing? *MIT Sloan Management Review*, 52(1): 41-49
- [14] Dowdell E B, Burgess A W, Flores J R. (2011).Online social networking patterns among adolescents, young adults, and sexual offenders. *American Journal of Nursing*, 111(7): 28-36
- [15] Cai Huajian, Lin Yongjia, Wu Qiuping, et al. (2008).Examining the measurement invariance between paper-and-pencil and Internet-administered tests in China. *Acta Psychologica Sinica*, 40(2): 228-239(in Chinese)
- [16] Bai Xinwen, Li Feng, Chen Yiwen. (2004).The measurement equivalence of social desirability. *Psychological Science*, 27(5): 1083-1086(in Chinese)
- [17] Fang Jiaming. (2009).Intention to participate in web surveys. Chengdu: University of Electronic Science and Technology of China(in Chinese)
- [18] Raju N S, Laffitte L J, Byrne B 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
- [19] Byrne B M, Shavelson R J, Muthen B. (1989).Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105(3): 456-466
- [20] Vandenberg R J, Lance C E. (2000).A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1): 4-70
- [21] Hurt H T, Joseph K, Cook C D. (1977).Scales for the measurement of innovativeness. *Human Communication Research*, 4(1): 58-65
- [22] Agarwal R, Parsad J. (1998).A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research*, 9(2): 204-215
- [23] Box G E P. (1949).A general distribution theory for a class of likelihood criteria. *Biometrika*, 36(3): 317-346
- [24] Chen F F. (2007).Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3): 464-504
- [25] Chan D, Schmitt N. (1997).Video-based versus paper-and pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82(1): 143-159
- [26] Scott J. (2000).*Social Network Analysis: a handbook*. London: Sage, 2nd edition, 87