TITLE: Assessing Measurement Invariance of the Survey of Perceived Organizational Support

(SPOS): Paper-and-Pencil vs. Online Administrations

AUTHORS: Özlü DOLMA,Serkan DOLMA

PAGES: 21-34

ORIGINAL PDF URL: https://dergipark.org.tr/tr/download/article-file/609104

Assessing Measurement Invariance of the Survey of Perceived Organizational Support (SPOS): Paper-and-Pencil vs. Online Administrations

Algılanan kurumsal destek ölçeğinin ölçüm değişmezliği: Kağıt-kalem ile online/web-tabanlı uygulamalarının karşılaştırılması

Özlü DOLMA [1], ozludolma@gmail.com

Serkan DOLMA [2], dolma@pau.edu.tr

Using multiple-group confirmatory factor analysis (CFA), this study investigates the measurement equivalence/measurement invariance of identical online/web-based and paper-and-pencil administrations of 8-item shortened version of the Survey of Perceived Organizational Support (SPOS) developed by Eisenberger, Huntington, Hutchison, and Sowa (1986). The results provided support for configural, metric, and scalar equivalence across administration modes, indicating that the psychometric properties of the 8-item shortened version of the survey are similar whether administered as a paper-and-pencil or an online measurement instrument.

Keywords: Measurement invariance, Measurement equivalence, Multiple-Group confirmatory factor analysis, Perceived organizational support

Jel Codes: C10, C80.

Bu araştırmada, çoklu grup doğrulayıcı faktör analizi kullanılarak Eisenberger, Huntington, Hutchison ve Sowa (1986) tarafından geliştirilmiş olan Algılanan Kurumsal Destek Ölçeğinin sekiz maddeli kısa versiyonunun online/web tabanlı ve kağıt–kalem ile uygulamaları bakımından ölçüm eşdeğerliliği/ölçüm değişmezliği test edilmiştir. Sonuçlar biçim, metrik ve skalar değişmezliğini desteklemiştir. Ölçeğin psikometrik özelliklerinin iki anket uygulama biçimi bakımından benzerlik gösterdiğine dair bulgular elde edilmiştir.

Anahtar Kelimeler: Ölçüm değişmezliği, Ölçüm eşdeğerliliği, Çoklu grup doğrulayıcı faktör analizi, Algılanan kurumsal destek

Jel Kodları: C10, C80.

---

[1] Ph. D. (Corresponding Author)

[2] Ph. D., Pamukkale University, Faculty of Economics and Business Administration

1. INTRODUCTION

The methods of data collection commonly used by researchers can be broadly classified into two main categories: (1) the interview surveys administered either face-to-face or via telephone and (2) self-administered surveys conducted by mail or by dropping off, via group administration, or using computer technology and online/Internet/web-based media (for details, see for example: Fowler, Jr., 2014, Gideon, 2012). Although, in past decades of survey research, mail, telephone, and face-to-face methods were the three dominant modes of data collection, with the introduction of computers and Internet to the survey process, the use of online/web-based surveys have become increasingly popular (Groves et al., 2009). Web-based administration is widely utilized as a data collection medium by researchers especially because responses can be gathered from large numbers of people in a very short amount of time at a fairly low cost (Dillman et al., 2014).

While researchers frequently use online measurement instruments to collect data, they usually overlook or take for granted the comparability of online instruments to their paper-and-pencil versions (Whitaker and McKinney, 2007). However, it is important to ensure that the measurement instrument used in a study functions the same across administration modes (i.e. measurement equivalence), especially if the researcher is planning to integrate data gathered via online surveys with those collected with paper-and-pencil based surveys. If a test of measurement invariance would reveal that the measurement instrument does not function in the same way across administration modes, it would not be appropriate to combine the data collected using these distinct methods (Fang et al., 2014; Teo, 2013; Whitaker and McKinney, 2007). Even if only online mode of data collection is employed through the research, it is still crucial to assure that the assumption of measurement invariance is not violated, so that one can justify the utilization of one specific mode of scale administration instead of another. That is, different administration modalities of a measurement instrument (e.g. paper-and-pencil vs online administration) should yield comparable scores within a respondent so that the researcher is allowed to use these different modes interchangeably (Hirai et al., 2011).

Our perusal of the relevant literature revealed that, especially in the last decade, several researchers used the online survey of POS as a measurement instrument. We have identified 22 studies where the online version of the SPOS was utilized. In all of these studies, researchers measured POS using a reduced number of items (e.g. 16-items, 8-items) and in half of them, scholars conducted factor analysis to confirm the unidimensional factor structure of the scale. Only in three of these studies, scholars reported CFA results in detail (e.g. fit indices, factor loadings). In most of these studies, researchers provided reliability evidence (Cronbach's alpha values ranged from .80 to .96) Although the results of these studies provided support for the psychometric adequacy of the online administration of shortened versions of SPOS, to the best of author's knowledge, no research to date has specifically investigated the psychometric properties of online-administered SPOS and the equivalence of its web-based survey versions with the analogous paper-and-pencil versions. This is the first study to examine different forms of measurement invariance (configural, metric, and scalar) across these two different modalities.

2. LITERATURE REVIEW

2.1. Survey of Perceived Organizational Support (SPOS)

Perceived organizational support (POS) is defined as employees' perception concerning the extent to which the organization values their contribution and cares about their well-being (Eisenberger et al., 1986: 501). The concept of POS has been of considerable importance in the organizational sciences literature for several decades for it is shown to be related to various employee attitudes, behaviors and outcomes of interest to researchers and practitioners. Meta-analytic studies reflect the amount of attention that POS has received in the literature (see, for example, Kurtessis et al., 2015; Ahmed and Nawaz, 2015; Riggle et al., 2009; or Rhoades and Eisenberger, 2002).

Previous research has shown that three general forms of perceived favorable treatment received from the organization, namely; fairness, supervisor support, and organizational rewards and job conditions have a substantial, positive effect on POS (Rhoades and Eisenberger (2002: 707). Similarly, findings from a recent meta-analysis by Kurtessis et al. (2015) have indicated that inspirational and supportive leadership, fair procedures, HR practices, and working conditions significantly contribute to POS. Research on the consequences of POS, on the other hand, has provided evidence that job satisfaction, organizational commitment, task and contextual performance, and intention to leave are among the most important work outcomes that POS would lead to (Riggle et al., 2009). Likewise, Kurtessis et al.'s (2015) meta-analytic review has shown that POS is related to positive attitudinal consequences such as affective commitment and organizational identification and desirable behavioral outcomes including increased in-role performance and organizational citizenship behavior (OCB) and reduced withdrawal behaviors.

Eisenberger et al. (1986) conceptualized POS as a unidimensional construct and developed 36 items to measure it. This measurement instrument has been termed the "Survey of Perceived of Organizational Support" (SPOS). Half of the items were worded negatively in order to control for the possible effects of agreement or disagreement response biases. Respondents were asked to indicate their level of agreement with SPOS items using a 7-point Likert-type scale with response options: 1 = 'strongly disagree', 2 = 'moderately disagree', 3 = 'slightly disagree', 4 = 'neither agree nor disagree', 5 = 'slightly agree', 6 = "moderately agree" and 7 = "strongly agree".

Eisenberger et al.'s (1986) scale development study consisted of two consecutive research. In the first study, they composed 36 statements to operationalize POS and analyzed the factor structure of the initial scale. In this study, the sample consisted of 361 employees from 9 different organizations. The reliability analysis of 36-item SPOS resulted in an alpha coefficient of .97 with item-total correlations ranging from .42 to .83. The principal component analysis revealed that all items load mainly on the presumptive perceived support factor which accounted for 48.3% of the total variance. In the second study, they examined the effect of employee's exchange ideology on the relationship between POS and absenteeism. For the second study, they used a shortened version of SPOS, including only 17 of the original SPOS items with highest factor loadings. Responses to the shortened SPOS and exchange-ideology questionnaire were subjected to separate factor analyses. The single factor solution for the 17-item SPOS accounted for 50% of the total variance. Then, by using

oblique rotation method for the combined set of responses to SPOS and exchange-ideology questionnaire, they obtained a two-factor solution, where items from each questionnaire formed their own factors. The correlation coefficient between these two concepts was -.10, providing evidence for their independence. Cronbach's alpha for SPOS was .93. Since in both of these studies, results supported the unidimensional factor structure of the survey and provided evidence for high internal reliability, scholars recommended the use of this shorter version of the survey. In the majority of studies on POS (probably, for practical reasons) researchers used shorter forms of the survey (e.g. 17-items, 8-items, and 3-items). In the present study, we used the 8-item shortened version of the SPOS.

2.2. Measurement Invariance in Organizational Research

Multiple-group confirmatory factor analysis (CFA) is a technique that can be used to examine the equivalence or invariance of the measurement across subgroups. This technique involves several successive steps in which a series of nested models are tested and compared. One of the basic ways of conducting multiple-group CFA is to build from the least constrained model (i.e. configural invariance or the equality of a number of factors and factor pattern matrices) to a fully constrained multiple-group CFA model (i.e. scalar invariance or the equality of item intercepts as well as factor loadings). At each step, a constraint of invariance (i.e. equal factor loadings) is added and the fit of the resulting model is evaluated to determine whether the added constraint has substantially worsened the fit of the model. Each extra step provides the opportunity to identify the parts of the model that are not equivalent across groups (Harrington, 2009).

A limited number of studies have investigated the invariance between paper-pencil and online administrations of measurement instruments used in organizational contexts.

Using multiple-group CFA, Cole et al. (2006) investigated the measurement equivalence of online and paper-and-pencil versions of the transformational leadership component of Bass and Avolio's (2000) Multifactor Leadership Questionnaire (TL-MLQ) in a sample of employees of a multinational organization located in 50 different countries. Moreover, they examined whether TL-MLQ will have equivalent relations with three theoretically related workplace constructs (collective efficacy, work-group cohesiveness, and collective goal commitment) across the two modes of scale administration. Their findings provided support for the configural, metric, scalar, measurement error, and relational equivalence between the paper-pencil and online administrations of the scale.

In another study, Deutskens et al. (2006) assessed the equivalence of online and mail surveys in a service quality context. Using multiple group CFA, they examined the measurement equivalence of a service quality survey across two groups of customers: (1) the traditional paper-and-pencil group which received a mail that contained an introduction letter, the questionnaire, and a prepaid return envelope and (2) the online group which received an email invitation, including a short introduction to the study with a request to participate and the hyperlink to the Web questionnaire. The survey was comprised of three subscales: (1) service call quality, (2) service visit quality, and (3) the intention to use the services of the provider again. The researchers recruited the participants for both groups were from a multinational office equipment manufacturer's customer database. They found that the means and variance-covariance matrices were equal across the online and offline surveys. In

other words, online and mail survey formats of these scales produced comparable results in terms of psychometric properties. Based on these findings they concluded that a mixed-mode survey method which combines the two survey modes (online and mail surveys) can be adopted as a data collection technique when measuring customers' service quality perceptions using these scales.

Whitaker and McKinney (2007) examined whether job satisfaction ratings will demonstrate measurement invariance over internet and paper-and-pencil administration modes in a sample of employed MBA and doctoral students. Although their initial results suggested that job satisfaction ratings are invariant across scale administration modes, when they tested for the effects of the respondents' demographic characteristics (gender and age) on latent job satisfaction variability across two modalities using multiple indicators, multiple causes models (MIMIC), their findings demonstrated that the respondent's age related to latent job satisfaction variability differently across paper-and-pencil and Internet administrations. In particular, they found that in the paper-and-pencil condition, job satisfaction ratings increased with respondent's age while in the Internet administration condition the responses to the job satisfaction scale were not related to age.

3. THE PURPOSE OF RESEARCH AND HYPOTHESIS

The present study addressed the comparability of scale administration modalities for the Survey of Perceived Organizational Support (SPOS: Eisenberger et al., 1986). This measurement instrument is commonly used in organizational studies to assess employees' general perceptions regarding the extent to which the organization appreciates their contributions and is concerned with their well-being. In addition to examining the psychometric properties of the SPOS for each measurement venue (paper-and-pencil based and online surveys), the current study investigated the factorial invariance across the measurement modalities using multiple-group CFA. Assessing measurement invariance is deemed essential to determine if the scale administration mode has an effect on participants' responses to the survey items. In order to gain confidence that the administration modalities can be used interchangeably, research is needed to ascertain that the psychometric properties of the scale are similar whether administered as a paper-and-pencil or an online measure.

Given the prevalence of research findings indicating that the psychometric properties of the online version of SPOS was promising and similar to those of its paper-and-paper versions, the current study tested the following scientific hypothesis: Responses to the 8-item shortened version of the SPOS will demonstrate measurement invariance over paper-and-pencil and online scale administration modes. This study will thus extend the findings of the existing literature by investigating whether different measurement modalities of the SPOS would provide comparable results.

4. METHOD

4.1. Participants

Online Sample. For the online survey, twenty retail stores of a prominent shoe retailing company operating in Turkey were recruited to take part in the study. These stores were located in ten different cities including Istanbul, Ankara, and Izmir. Links to the online surveys were sent to 565 sales representatives working in these stores. Of those who received

the link, 337 completed the survey, which represents a response rate of 59.6%. In order to encourage honest and candid responses and to allow for anonymity, we did not collect any identifying demographic information such as gender, age, or status, except the organizational tenure. The majority of the participants had an organizational tenure of five years or less (92.9%). Tenure with the organization of 6.8% of the participants ranged between six to ten years. Only 0.3% of the participants' organizational tenure was more than ten years.

Paper-and-Pencil Sample. In the paper-and-pencil sample, participants were 315 white-collar employees from four information technology companies located in Istanbul and Denizli, Turkey (55.6% of the participants) and from the headquarters of the aforementioned shoe retailing company from which we collected the online data. Most of the participants (71.1%) had worked for their current company for five or less than five years. Tenure with the organization of 16.2% of the participants ranged between six to ten years. 8.6% of the participants' organizational tenure was more than ten years. 4.1% of the participants did not indicate their organizational tenure.

4.2. Measure

Since this study investigated the measurement invariance of online and paper-and-pencil based administrations of the 8-item shortened version of the Survey of Perceived Organizational Support (SPOS) developed by Eisenberger et al. (1986), we asked participants to complete the items of this survey. Items employed in the current study, from no. 1 to 8, correspond to the items no. 1, 9, 27, 21, 4, 8, 25, 35 of the original 36-item scale developed by Eisenberger et al. (1986: 502), respectively. Respondents provided their level of agreement to the all positively worded items on a 5-point Likert-type rating scale with response options ranging from 1=strongly disagree to 5=strongly agree.

Reliability analysis resulted in a coefficient alpha of .94 with corrected item-total correlations ranging from .67 to .83 for the paper-and-pencil based survey and a coefficient alpha of .91 with item-total correlations ranging from .60 to .79 for the online survey. Composite reliability coefficients for paper-and-pencil based and the online surveys were .94 and .91, respectively.

4.3. Procedure

Surveys were administered in two different assessment settings: (1) Online survey and (2) Paper-and-pencil based survey.

For the online group, data were collected as a part of a larger research project on work teams and thus the survey participants completed included also measures of other variables pertaining to that research project. Before the links to the online survey were sent to retail store computers, each store manager was informed about the study and the online survey to be completed by the sales representatives. At the beginning of the online survey, an introduction page was presented to participants which was linked to the survey page and included brief information about the study and the instruction for the online survey. Each participant completed the survey alone in an isolated room in the store during the work shift. Each item was displayed on separate webpage and respondents were not allowed to

skip items. That is, in order to complete the survey participants had to provide a response to each of the items.

For the online survey, data were collected specifically for this study and the survey form included only the SPOS. Individuals in the paper-and-pencil group received a survey form, which included a short introduction to the study with a request to participate with, in a return envelope. Those who volunteered to participate in the study, completed the survey in their offices within working hours. In the online group, items were presented to participants via Surveey® online survey system. The paper-and-pencil form and the online version of the SPOS were identical in response format, item order, and length.

4.4. Data Analysis

Descriptive statistics and alpha coefficients were calculated using IBM SPSS Statistics 20.0. The series of invariance tests for both the measurement model and then on the structural model were performed utilizing Mplus 7.0 (Muthén & Muthén, 1998–2015). Composite reliability coefficients were calculated using Mplus 7.0.

Prior to the analyses, paper-and-pencil based surveys were examined to determine whether there were discernible patterns of nonresponse. This examination revealed that the surveys had missing values at random. Only ten of the paper-and-pencil based surveys included missing data. Of these surveys, one had non-response for two items. There was only one omitted item in each of the remaining nine surveys. Since in the online group respondents were not allowed to skip any of the items, there was no missing data in the online group's data set.

For the percentage of missing data was small (0.44%), following Mellenbergh's (2011) recommendation, we imputed the missing data using subject mean substitution. The missing observations of participants were replaced with the response category that is closest to the mean of their non-missing responses.

Distributions of item scores were also inspected to determine whether there were significant departures from univariate normality. When we checked the skewness and kurtosis values of both the online and the paper-and-pencil data sets, we observed only minor deviations from normality. Although none of the items' skewness and kurtosis values were extreme and varied approximately between -1.0 and +1.0 in both groups, to err on the side of caution, instead of the standard Maximum Likelihood (ML) estimation method which is utilized for data sets that are multivariate normal, we used the Satorra-Bentler scaled chi-square statistic (χ^2_{SB}) with robust standard errors and corrected model test statistics (Satorra and Bentler, 1994) for assessing the fit of the CFA models. We followed Cheung and Rensvold's (2002) and Chen's (2007) guidelines in interpreting the change in model fit statistics. The rationale for inspecting the changes in approximate fit indices for testing measurement invariance is discussed later in detail in the results section.

5. RESULTS

Before the multiple-group CFA invariance evaluation, we tested the one-factor structure of POS separately in each group. Model fit indices indicated that the one-factor model fit the data quite well for both the online (χ^2_{SB}=19,430, df=20, *p*=.4941, CFI=1.000, TLI=1.001, SRMR=.020, RMSEA=.000) and the paper-and-pencil based data (χ^2_{SB}=55,637, df=20, *p*<.001,

CFI=.976, TLI=.967, SRMR=.026, RMSEA=.075). In both groups, all estimated factor loadings were statistically significant (*p*<.001) and the standardized factor loadings ranged from .627 to .837 for the online group and from .701 to .860 for the paper-and-pencil based group.

Since separate factor analyses revealed that the measurement models are identical across the two groups, in the next step, using the pooled data set, we conducted the simultaneous analysis of equal form to determine a baseline model against which we were going to test the fit of a more restrictive model of equal factor loadings. In testing this model, no equality constraints were imposed and factor loadings, measurement intercepts, and error variances were freely estimated. As in single group SEM models, for both groups factor means were fixed at zero for model identification purposes. The configural one-factor model provided a reasonably good fit to the data (χ^2_{SB}=72,023, df=40, *p*=.0014, CFI=.987, TLI=.982, SRMR=.023, RMSEA=.05), indicating that equal factor form (i.e. configural invariance) held across the two groups. Results of the invariance tests are presented in Table 1.

After determining the baseline model, we followed a procedure in which we tested the fit of a series of increasingly stringent models. Traditionally, the chi-square difference test between nested models is used to determine the measurement invariance (Brown, 2015). However, the chi-square statistic is known to be sensitive to sample size (Cheung and Rensvold, 2002; Meade et al. 2006). Given that the chi-square differences also have a chi-square distribution, they are also sample size dependent (Brannick, 1995). Thus, especially in very large samples, the chi-square difference test could be statistically significant (i.e. lack of measurement invariance) even though the absolute differences in parameter estimates indicate little difference in fit (Cheung and Rensvold, 2002; Kline, 2016; Meade et al., 2008). To circumvent this problem, as recommended by Little (2013), rather than relying on the results of the chi-square difference tests, we inspected the changes in approximate fit indices when testing the measurement invariance.

The findings of three prominent Monte Carlo simulation studies (i.e. Cheung & Rensvold's [2002], Chen [2007], and Meade et al.'s [2008]) also provided support for the recommendation that the changes in approximate fit indices can be examined to establish measurement invariance. In these three studies, scholars suggested different cutoff criteria for changes in fit indices to evaluate invariance. Cheung and Rensvold's (2002) suggested that the change in CFI value which is less than or equal to .01 indicates that the null hypothesis postulating stricter invariance should not be rejected. Chen (2007) also recommended that when group sizes are equal and larger than 300, a change in CFI larger than .01, supplemented by a change in RMSEA larger than .015, can work reasonably well in detecting the lack of invariance. On the other hand, based on their simulation study Meade et al.'s (2008) contended that the recommended value for delta CFI of Cheung and Rensvold (2002) seems excessively large and advised that the change in CFI value that exceed .002 should be used as a criterion for evidence of non-invariance. In this study, we employed Cheung and Rensvold's (2002) and Chen's (2007) cutoff criteria for invariance testing.

Next, we evaluated whether the factor loadings of the SPOS items are equivalent in both groups (i.e. metric invariance). We ran a model where only the factor loadings were constrained equal across groups but the intercepts and error variance were allowed to differ between groups. Again, factor means were fixed at zero for both groups. Comparison of this

model with the configural model yielded a $\Delta\chi^2_{SB}$ of 13,414, which was only marginally statistically significant (Δdf= 7, *p*=.04). Moreover, in keeping with Cheung and Rensvold's (2002) and Chen's (2007) guidelines, when we inspected the changes in approximate fit indices (CFA and RMSEA) and compared them to the recommended cutoff criteria, the results indicated that invariance hypothesis should not be rejected (ΔCFI = .002 and ΔRMSEA<.001) providing support for metric invariance.

In the next step, we examined whether the item intercepts were also equal across assessment modalities (i.e. scalar invariance). To this aim, we constrained the item intercepts to be equal across groups in addition to the constraints imposed in the previous step while fixing the factor mean at zero in the first group and freely estimating it in the second group. The difference in model fit between this model and the metric invariance model was statistically significant ($\Delta\chi^2_{SB}$ = 21,911, Δdf=7, *p*<.05). However, the examination of the changes in approximate fit indices revealed that CFA and RMSEA were less than the critical thresholds as recommended by Cheung and Rensvold (2002) and Chen (2007) (ΔCFI =.006 and ΔRMSEA=.005) supporting the scalar invariance hypothesis.

After we obtained metric and scalar invariance, we also performed the test of equality of error variances across the groups (i.e. strict invariance) although it is considered an overly restrictive test by many scholars (e.g. Little, 2013; Byrne, 2016). Some researchers contented that the equality of the indicator error variances is not a substantively important condition of measurement invariance and is also not a prerequisite of subsequent invariance testing (Brown, 2015). Moreover, according to some scholars (e.g., Little, 2013) enforcing the equality of error variances can be problematic to invariance evaluation because the ill fit introduced by these constraints may bias the remaining parameter estimates (e.g., factor loadings, factor variances) (Brown, 2015: 262). Nevertheless, we tested for the equality of between-group error variance for completeness. Applying this additional constraint of equal error variances resulted in a $\Delta\chi^2_{SB}$ of 211,981, which was statistically significant (Δdf=8, p<.05). Changes in approximate fit indexes also revealed that there was a salient deterioration in model fit (ΔCFI =.08 and ΔRMSEA=.058). Consequently, we rejected the null hypothesis stating that error variance is invariant across groups. That is, for each scale item, the portion of item variance which is not attributable to the variance of the latent variable POS was not equal across the paper-and-pencil and online groups.

After measurement invariance is achieved to a certain level, if making inferences across populations is of researchers' substantive interest, structural invariance testing can be conducted. In this study, the primary focus was to test for the invariance of the measurement model across the paper-and-pencil and online forms of SPOS. Since the populations differed not only in terms of the survey administration modalities but also in terms of various other personal characteristics (e.g. employing organization, occupation, position), neither comparing latent means and factor variances of the models across the two groups was of substantive importance, nor lack of latent mean invariance was easily interpretable. Moreover, a meaningful interpretation of group mean contrasts is impracticable. Nevertheless, we examined the equivalence of structural parameters across online and paper-and-pencil based groups for exploratory purposes. First, we constrained the latent factor variances to be equal across groups in addition to the constraints imposed in the scalar invariance testing, while fixing the latent mean to zero in one group and freely estimating

the latent mean in the second group. Adding this constraint resulted in a $\Delta\chi^2_{SB}$ of 6,090, which was statistically significant (Δdf=1, p<.05). However, the changes in approximate fit indexes revealed that the equal factor variances model did not result in a remarkable degradation of fit relative to the equal item intercepts model (ΔCFI=.002 and ΔRMSEA=.002). Since our study entailed a one-factor measurement model, the invariance evaluation of factor covariances was not applicable. Finally, the latent mean of the online group was held equal to the paper-and-pencil based group to test the equality of factor means across groups. This constraint deteriorated the fit of the model substantially ($\Delta\chi^2_{SB}$=39,512, Δdf=1, p<.01, ΔCFI=.015, ΔRMSEA=.016). Although it is not substantively very meaningful, based on these results it can be concluded that the two groups differed significantly in their latent mean POS scores.

Table 1. Results of the measurement invariance tests

Invariance	χ^2_{SB}	df	SCF	CFI	RMSEA	SRMR	Model Difference Tests				
							$\Delta\chi^2_{SB}$	Δdf	p	ΔCFI	ΔRMSEA
Measurement											
Configural	72,515	40	1.392	.987	.050	.023					
Metric	85,929	47	1.321	.985	.050	.041	13,414	7	.040	.002	0
Scalar	107,840	54	1.285	.979	.055	.047	21,911	7	.004	.006	.005
Strict	319,821	62	1.309	.899	.113	.078	211,981	8	<.001	.080	.058
Structural											
Equal Factor Variances	113,930	55	1.272	.977	.057	.097	6,090	1	.001	.002	.002
Equal Factor Variances & Means	153,442	56	1.261	.962	.073	.158	39,512	1	<.001	.015	.016

Note. χ^2_{SB} = Satorra-Bentler Scaled Chi-square Statistic; SCF = Scaling Correction Factor; CFI = Comparative Fit Index; RMSEA = Root Mean Squared Error of Approximation; CI = Confidence Interval; SRMR = Standardized Root Mean Squared Residual.

6. DISCUSSION

The issue of measurement invariance represents a matter of fundamental importance, as researchers increasingly rely solely or partially on the online versions of the psychological measurement instruments to collect data. Although delivering surveys over the Internet may offer significant cost advantages, researcher should be attentive to potential effects peculiar to this scale administration method, especially when the scale was originally developed in another format (e.g. paper-and-pencil form). That is, measurement equivalence of various data collection methods may not be taken for granted and has to be tested. As Bartram (2006: 33) cautioned, "when a test is presented in some medium other than the one in which it was developed, it is necessary to check the equivalence of the new form". Since the SPOS was developed as a paper-and-pencil survey, it is crucial to investigate the equivalence of the original paper-and-pencil based version to the online version, in order to make sure that the data obtained via both assessment methods are comparable.

The current study was the first to examine the comparability of the psychometric properties of the SPOS administered under two different conditions: the conventional paper-and-pencil based measurement and the increasingly popular mode of online survey administration via Internet. The initial CFA results demonstrated that the SPOS taps the same underlying

unidimensional factor structure when administered online as well as on paper. In addition, the reliability of the SPOS was adequate within each administration mode. The findings of measurement equivalence tests revealed that the data obtained from the two measurement modalities were comparable. Results supported the configural, metric, and scalar invariance of the SPOS across the online and paper-and-pencil conditions.

Structural invariance tests were conducted for exploratory purposes only and the results of these tests supported the equality of factor variances across the two modes, whereas they failed to provide evidence for the equivalence of factor means. However, as mentioned before, for the present study, the lack of factor means equivalence cannot be easily interpreted, since the populations differed with regard to variables other than the membership to scale administration groups that may provide alternative explanations to the observed difference between the two groups.

A potential limitation of the study arises from the fact that we were not able to use a truly random process to sort participants into two assessment groups. The fact that companies were reluctant to share their list of employees and their contact information, and other logistic and practical difficulties precluded us from carrying out random assignment procedures. Therefore, instead of randomly assigning each participant to one of the test administration modes, a method that is more palatable to the administration was followed, in which we applied the same administration mode to all participating employees of a particular company. This convenience of using intact groups, of course, comes with a price: As a consequence, we were not able to rule out the effects of the unmeasured characteristics of the sample subjects, which may account for the group differences (e.g. unequal latent means). Future research employing random assignment of research participants from a common population to the groups of paper-and-pencil based and online measurement settings would provide more rigorous information about the measurement and structural invariance between the two modalities.

In this study, participants completed either the paper-pencil or online version of the survey. Future research adopting a repeated measures design, where each participant would complete both versions of the SPOS at two different time points over a specific period, may also provide evidence of potential within-individual differences across online and paper-and-pencil conditions.

Furthermore, in current study, online surveys were administered using computers. Future research can investigate the comparability of the data obtained via paper-and-pencil based survey and the online tools other than computers such as smart phones and tablets. When findings from the present research are replicated in such studies, conclusions can be drawn regarding the consistency of the psychometric properties of the SPOS across various online assessment tools and the paper-and-pencil formats and researchers may be able to use the different types of online tools interchangeably or within the same study.

In sum, the results of the present study provided initial evidence that the web-based version of the survey is a reliable and valid means for assessing POS and the online and paper-and-pencil based administrations of the SPOS are comparable. However, considering the limitations of the study, further efforts are warranted to replicate the comparisons between

paper-and-pencil and online administration modalities of the SPOS with regard to measurement and structural parameters.

REFERENCES

Ahmed, I., & Nawaz, M. M. (2015). Antecedents and Outcomes of Perceived Organizational Support: A Literature Survey Approach. *Journal of Management Development, 34*(7), 867-880.

Bartram, D. (2005). *Testing On the Internet: Issues, Challenges and Opportunities in the Field of Occupational Assessment*. In D. Bartram & R. Hambleton (Eds.), Computer-Based Testing and The Internet: Issues and Advances (pp. 13-37). New York: Wiley.

Bass, B. M., & Avolio, B. J. (2000). Multifactor Leadership Questionnaire: Technical Report, Leader Form, Rater Form, And Scoring Key For MLQ Form 5x-Short. Redwood City, CA: Mindgarden.

Brannick, M. T. (1995). Critical Comments on Applying Covariance Structure Modeling. *Journal of Organizational Behavior, 16*, 201-213.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press.

Byrne, B. M. (2016). *Structural Equation Modeling With AMOS: Basic Concepts, Applications and Programming*. New York, NY: Routledge/Taylor & Francis.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling, 14*, 464-504.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-Of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling, 9*, 233–255.

Cole, M. S., Bedeian, A. G., & Field, H. S. (2006). The Measurement Equivalence of Web-Based and Paper-And-Pencil Measures of Transformational Leadership. *Organizational Research Methods, 9*, 339-368.

Deutskens, E., de Ruyter, K., & Wetzels, M. (2006). An Assessment of Equivalence between Online and Mail Surveys in Service Research. *Journal of Service Research, 8*, 346-355.

Dillman, D. A., Smyth, J. D., & Christin, L. M. (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: Wiley & Sons.

Eisenberger, R., Hungtington, R., Hutchison, S., & Sowa, D. (1986). Perceived Organizational Support. *Journal of Applied Psychology, 71*, 500-507.

Fang, J., Wen, C., & Prybutok V. (2013). The Equivalence of Internet versus Paper-Based Surveys in It/Is Adoption Research in Collectivistic Cultures: The Impact of Satisficing. *Behaviour & Information Technology, 32*(5), 480-490.

Fowler, F.J. (2014). *Survey Research Methods*. Thousand Oaks, CA: Sage.

Gideon, E. (2012). *Handbook of Survey Methodology for the Social Sciences*. New York, NY: Springer.

Groves, R. M., Fowler, F. J., Couper, M. K., Ledowski, J. M., & Singer, E. (2009). *Survey Methodology*. Hoboken, NJ: John Wiley.

Harrington, D. (2009). *Confirmatory Factor Analysis*. New York, NY: Oxford University Press.

Hirai, M., Vernon, L., Clum, G., & Skidmore, S. (2011). Psychometric Properties and Administration Measurement Invariance of Social Phobia Symptom Measures: Paper-Pencil vs. Internet Administrations. *Journal of Psychopathology and Behavioral Assessment, 33*, 470-479.

Kline, R. B. (2016). Principles and Practice of Structural Equation Modeling. New York, NY: Guilford Press.

Kurtessis, J. N., Eisenberger, R., Ford, M. T., Buffardi, L. C., Stewart, K. A., & Adis, C. S. (2015). Perceived Organizational Support: A Meta-Analytic Evaluation of Organizational Support Theory. *Journal of Management, 20*(10), 1-31.

Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. New York, NY: Guilford Press.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2006, August). *The Utility of Alternative Fit Indices in Tests of Measurement Invariance.* Paper presented at the Annual Academy of Management Conference, Atlanta, GA.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and Sensitivity of Alternative Fit Indices in Tests of Measurement Invariance. *Journal of Applied Psychology, 93*, 568-592.

Mellenbergh, G. J. (2011). *A Conceptual Introduction to Psychometrics: Development, Analysis, and Application of Psychological and Educational Tests*. The Hague, Netherlands: Eleven International Publishing.

Muthén, L.K., & Muthén, B.O. (1998-2015). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.

Rhoades, L., & Eisenberger, R. (2002). Perceived Organizational Support: A Review of the Literature. *Journal of Applied Psychology, 87*, 698-714.

Riggle, R. J., Edmondson, D. R., & Hansen, J. D. (2009). A Meta-Analysis of the Relationship between Perceived Organizational Support and Job Outcomes: 20 Years of Research. *Journal of Business Research, 62*, 1027-1030.

Satorra, A., & Bentler, P. M. (1994). *Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis*. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA, US: Sage Publications, Inc.

Teo, T. (2013). Online and Paper-Based Survey Data: Are They Equivalent? *British Journal of Educational Technology, 44*(6), E196-E198.

Whitaker, B. G., & McKinney, J. L. (2007). Assessing the Measurement Invariance of Latent Job Satisfaction Ratings across Survey Administration Modes for Respondent Subgroups: A Mimic Modeling Approach. *Behavior Research Methods, 39*, 502-509.