OmniaScience

Intangible Capital

IC, 2017 – 13(5): 879-901 – Online ISSN: 1697-9818 – Print ISSN: 2014-3214

https://doi.org/10.3926/ic.1049

Testing invariance between web and paper students satisfaction surveys. A case study

Monica Martinez-Gomez[1], Juan A. Marin-Garcia[2], Martha Giraldo O'Meara[3]

[1]*Departamento de Estadística e Investigación Operativa. Universidad Politécnica de Valencia (Spain)*

[2]*ROGLE. Departamento de Organización de Empresas. Universitat Politècnica de València (Spain)*

[3]*Universitat de Valencia (Spain)*

momargo@eio.upv.es, jamarin@omp.upv.es, marthaomeara@hotmail.com

Received April, 2017

Accepted October, 2017

Abstract

Purpose: This paper studied the measurement invariance (MI) across web-based and paper-based surveys to evidence if both techniques of data collection can be regarded as equivalent.

Design/methodology: We develop a multigroup confirmatory factor analysis (MGCFA) with Maximum Likelihood Estimation to asses measurement invariance of the Job Diagnostic Survey (JDS) adapted to teaching, with data collected from paper and web surveys. Sample from paper surveys was constituted by 294 student of a Spanish public university in the academic years 2007-08, 2008-09 and 2009-10. Internet surveys were administered through an open source survey application called Lime Survey. We received 241 completed questionnaires.

Findings: Results show that metric invariance, covariance invariance, variance of latent factors invariance and measurement errors invariance can be established between two groups. We can conclude that both methods of collecting data can be considered equivalent.

Research limitations/implications: This study was done with a particular sample and strict focus questionnaire and we might not generalize the findings. It should be extended in the future to include other universities and graduate students.

Originality/value: Results showed that the factor structures remained invariant across the internet-based and paper-based groups, that is to say, both methods of collecting data can be considered equivalent, with the same factor structure, factor loadings, measurement errors of factors and the same reliability. These findings are useful for researchers since they add a new sample in which web and paper questionnaires are equivalent and for teachers to desire to change the teaching methodology at university, encourage students' participation and teamwork through active methodologies.

Keywords: Measurement equivalence, Students' satisfaction and motivation, Job Diagnostic Survey, Multigroup confirmatory analysis, Higher education, Active methodologies

Jel Codes: I23, C38, M10

1. Introduction

Surveys are particularly significant in education and science researches. In the past, surveys were always provided in paper, but in the last few years and since internet has turned into a powerful and efficient tool for searching and collecting information, the trend is to use online surveys.

However, different studies (e.i. Aster, 2004; Cook, Heath & Thompson, 2000; Hogg, 2003; Nulty, 2008) suggested that in many situations it is not possible to apply only one mode of collecting data and proposed to use a mixed-mode design as a solution to increase the level of response.

Many researchers have seemingly assumed that paper and web surveys exhibit adequate cross-mode equivalence, but when integrating data collected from internet surveys with traditional paper-based surveys, researchers must ensure about the reliability, validity and comparability of data collected (Vandenberg & Lance 2000). That is to say, to evidence the measurement equivalence (MI) of these two modes of surveys and that the measured latent construct has the same theoretical pattern (Cole, Bedeian & Feild, 2006; Miles & King, 1998).

The establishment of measurement invariance across groups is a prerequiste to work data collects across different groups and researchers have pointed out that is necessary to ensure measurement equivalence in each organisational research (i.e.: Jöreskog, 1971; Byrne, 1989, Elosua, 2005; Vandenberg & Lance 2000).

Therefore, we need to study the invariance of the psychometric properties of both modes to collect data.

Nowadays, measurement invariance is often tested with a multiple-group confirmatory factor analysis (CFA), in the framework of structural equation models, as suggest several studies: Vandenberg and Lance (2000), Chen, Sousa and West (2005), Cheung and Rensvold (2002) and Cheung (2008).

The main purpose of this study is to evaluate whether the data collected through web and paper-based survey can be regarded as equivalent. In order to test for measurement invariance across these two survey modes, confirmatory factor analysis (CFA) was used to evaluate to MI of the Job Diagnostic Survey (JDS) adapted to teaching.

The rest of the paper is structured as follows. First, the theoretical framework of advantages and weakness of internet, paper and mix-mode surveys is presented. Second, the background of MI between paper and web surveys is described, summed up later on students' satisfaction surveys. Then, we summarize the research methodology and the main results we conclude with the discussion of the main findings achieved in our analysis.

2. Theoretical framework

2.1. Internet, paper-based and mixed-mode surveys

The proliferation of online surveys has generated over the last years, several reviews about the strengths and limitations of web surveys compared to the traditional paper-based surveys. However findings in many of these studies were contradictory. The main conclusions of some of these studies, highlight by De Beuckelaer and Lievens (2009) are summarize at Table 1.

In spite of the proliferation of web surveys, some studies reveal that many times people, especially students, prefer to answer in paper format. For example, Van Gelder, Bretveld and Roeleveld (2010) in a study with young students found that the 83% of them preferred to respond in a paper mode. This same trend has been identified by Hohwü, Lyshol, Gissler, Jonsson, Petzold and Obel (2013), in a study with Danish students.

In order to improve the response rate, the trend of many works is to combine internet surveys with a more conventional mode of data collection, such as, paper surveys (Yun & Trumbo, 2000). This trend is known as mixed-mode surveys.

Advantages of web surveys	Disadvantages of web surveys
Web-based surveys are completed by a larger amount of people, at a lower cost than paper-based surveys and data processing is more efficient (e.i. Schonlau, Fricker & Elliott, 2002). Data collection and processing are immediate, error rate is lower because data are not entered manually, fewer human resources services are required, costs are lower than in paper-based surveys and data analysis and achievement of findings is faster (e.i. Dillman Smyth & Christian, 2009; Martins, 2010). Less costly (e.i. Dillman, 2000; Kraut & Saari, 1999; Schaeffer & Dillman, 1998; Sproull, 1986; Young, Daum, Robie & Macey, 2000; Yun & Trumbo, 2000). Lead to faster survey responses (e.i. Schaeffer & Dillman, 1998; Sproull, 1986). Allow greater flexibility in survey design (Dillman, 2000). Offer a wider variety of response formats (Simsek & Veiga, 2001). Wider geographical reach (Epstein, Klinkenberg, Wiley & McKinley, 2001). There is no human (coding) errors (e.i. Cook et al. 2000; Roberts, Konczak & Macan, 2004). Are free of experimenter bias (e.i. Reips, 2000). Are less sensitive to order of question effects due to the ease of randomising questions (Bowling, 2005). Not have many missing values (Stanton, 1998). Greater flexibility in survey design and wider variety of response formats (Simsek & Veiga, 2001). Wider geographical reach (Epstein et al., 2001).	Web surveys requires a computer and internet access, respondents require online experience, technological variations, the confidentiality of the responses is reduced due to the nature of the ID systems used, system errors or server problems, low response rate and surveys are impersonal because there is no human contact (e.i. Dillman et al., 2009; Martins, 2010) (ID Systems= Autoidentyfication Systems). The use of internet to collect data is restricted by coverage limitations and the lack of willingness of people to respond for different reasons (e.i. Fang, Wen & Prybutok, 2014; Bosnjak, Tuten & Wittmann, 2005; Fan & Yan, 2010; Fang, Wen, & Pavur, 2012; Göritz, 2006). Higher non-response rates e.i. (Schaeffer & Dillman, 1998; Sproull, 1986). Higher probability of getting dishonest answers (Lautenschlager & Flaherty, 1990). Potential technological problems (Kraut & Saari, 1999). Decreased item reliability due to somewhat higher measurement errors (Stanton, 1998). Possibility of multiple submissions (Reips, 2000).

Table 1. Strengths and limitations of web and paper surveys

2.2. Measurement invariance of web and paper-based surveys

Nowadays, to aggregate data collected from internet and paper-based surveys, researchers must ensure that both survey modes are comparable. For this, it is necessary to check the MI between the two different survey modes.

There are previous studies in different areas, about if the findings achieved with paper and online questionnaires are equivalent, since the intention of many researchers is to increase the number of responses, using both jointly (mixed-model) (i.e. King & Miles, 1995; Fouladi, McCarthy & Moller, 2002; Meade, Michels & Lautenschlager, 2007; Steinmetz, Schmidt, Tina-Booh, Wieczorek & Schwartz, 2009; Yu & Yu, 2007).

The statistical techniques and fields used in such studies are very diverse. Initally all of these studies used traditional techniques to analyse the equivalence between different groups. For exa mple, Riva, Teruzzi and Anolli (2003) used an exploratory factor analysis (EFA) to reveal that web or paper surveys show equivalent levels of reliability, extracted number of factors and factor loadings. Buchanan and Smith (1999) carried out an exploratory factor analysis and a multigroup confirmatory analysis on the reviewed version of Gangestad and Snyder's (1985) self-control questionnaire, in order to analyse if there were differences between paper and web questionnaires. The findings they achieved were essentially three: psychometric properties were favoured when students completed an online surveys, the factor structure of the questionnaire was invariant in both formats and the honesty of students is higher when responding via web.

Along the same line, Herrero and Meneses (2006), carried out a study on the reduced versions of Perceived Stress Scale questionnaires (Cohen, Kamarck & Mermelstein, 1983) and the Center for Epidemiology Studies-Depression Scale CES-D (Radloff, 1977), achiev ing acceptable values of internal consistence through α-Cronbach, which revealed that both structures were invariant regardless of the format used and that paper and web surveys were virtually equivalent.

However, these findings do not guarantee the invariance of psychological properties of Internet-based and paper-based instruments, since EFA is a sample-dependent technique and no criterion exists for comparing differences in the factor analysis parametres based on different groups. (Sen-Chi & Min-Ning, 2007). Later, Walt, Atwood and Mann (2008), tested whether or not survey medium, electronic or paper format, had a significant effect on the results achieved, reliability, item mean, response rate, response completeness, and factor analysis comparisons across survey media. However they didn't use confirmatory factor analysis (CFA) that nowadays is the more common method of comparing invariance between two groups. So, as most previous studies that compare online and mail surveys, it has methodological limitations.

Afterwards, De Beuckelaer and Lievens (2009), examined the measurement equivalence between internet data collection and the traditional paper-pencil method with a organisational survey in 16 countries. In that paper they made an over review of prior studies testing the equivalence invariance across multiple methods of data collection and the main relevant conclusions of them. The found that scalar invariance between internet and paper-pencil surveys was fulfill across the countries.

Finally, some studies using confirmatory factor analysis to assess invariance between paper and web surveys included the levels of configural, metric, scalar, covariance invariance, means variance invariance and variance of latent variables, for example Fang et al. (2014), Davidov and De pner (2011) or Leung and Kember (2005), but results show contradictory findings.

The lack of consistency in the results, produces an important area of research. Fang et al. (2014) recommending that when we conducting research in collecting data from distinct survey modes, we should concern themselves with the measurement invariance across survey modes.

2.3. Purpose and contributions of present study

The purpose of this study is to answer the research question: "what differences existing responses between paper and web-based survey methods?" We examine the measurement equivalence across data collection modes surveys with data collected from the Job Diagnostic Survey (JDS) adapted to teaching (Giraldo-O'Meara, Marin-Garcia & Martínez-Gómez, 2014). This questionnaire was developed to check if active learning improve students' satisfaction and motivation, according with many authors (i.e. Aydin & Ceylan, 2008; Barak, Ben-Chaim & Zoller, 2007; Ebenezer, Columbus, Kaya, Zhang & Ebenezer, 2012; Ismail, Mashkuri, Sulaiman & Kee Hock, 2011; Marbach-Ad & Sokolove, 2002; Orgambídez-Ramos, Borregó-Alés & Mendoza-Sierra, 2014).

This study contributes to the existing literature on survey research in some ways. On the one hand, results provide researchers, an assessment of equivalence between Internet and paper-based surveys and information on the feasibility of integrating data collected via Internet surveys by offering empirical evidence using data collected from JDS. On the other hand, to evaluate if active methodologies can promote higher motivation and satisfaction on students (Trullas & Enache, 2011).

3. Methodology

3.1. Sample

The total sample was constituted by 535 student of a Spanish public university in the academic years 2007-08, 2008-09 and 2009-10. 294 questionnaires were completed in paper-based survey in the classroom, 10 minutes before the end of the lesson. Internet surveys were administered through an open source survey application called Lime Survey. We received 241 completed questionnaires. This sample was used in other study (Martínez Gómez, Marin-Garcia & Giraldo-O'Meara, 2016)

3.2. Instrument

In the present study, we used the validated version of Job Diagnostic Survey (JDS) (Hackman & Oldham, 1975), adapted teaching (Martínez-Gómez & Marín-García, 2009; Giraldo-O'Meara et al., 2014; Martínez Gómez et al., 2016) to test invariance between paper and web based surveys. The JDS (Hackman & Oldham, 1975, 1976, 1980) is one of the main tools to evaluate how estimulating a job position. As we datailed in previous studies (Giraldo-O'Meara et al., 2014; Martínez Gómez et al., 2016), its adapted version includes a satisfaction single-item scale (SAT), a motivating poten tial score (MPS) and the job characteristics scales (Figure 1).

4. Method of analysis

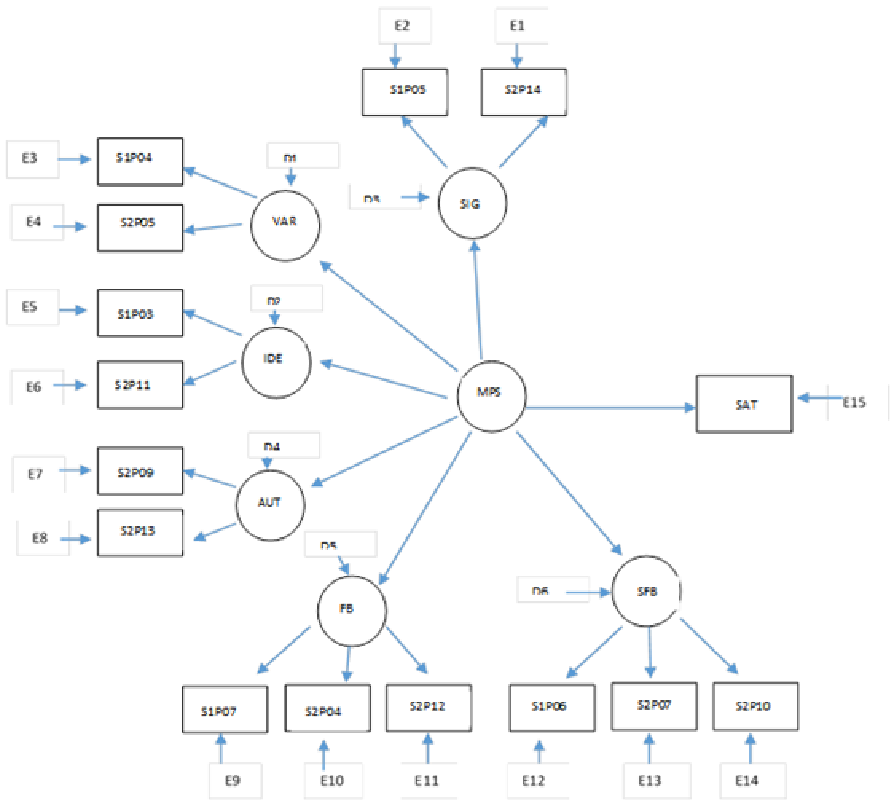
First of all, a variance analysis (ANOVA) was performed, to see if there were significant differences in the means for the items between both groups.

Consistent with previous studies, (Giraldo-O'Meara et al., 2014) where the model factor structure had been validated in the total sample, we examined the reliability scale of both samples, paper and web surveys, separately. For that purpose, compound reliability (CR =cut-off value .7) and extracted variance (EV= cut-off value .5) (Hair, Anderson, Thatam & Black, 1995) were used as measurements. We also checked the squared correlations coefficient of items and the goodness of fit indexes of the confirmatory factor analysis.

To assessment of model fit, apart from traditional fit indices (Chi-square), we relied on other measures of model fit (Bollen & Long, 1993; Brown & Cudek, 1993; Santos-Rego, Godás-Otero, Lorenzo-Moledo & Gómez Fraguela, 2010). In particular, we used: the Comparative Fit Index (Bentler & Bonett, 1980), the Tucker-Lewis Index (TLI), which is also referred to as the Bentler-Bonett Non-Normed Fit Index (NNFI; Bentler & Bonett, 1980), the Root Mean Square Error of Approximation

(RMSEA) (Steiger, 1990; Ullman & Bentler, 2004), and the Standardized Root Mean Square Residual (SRMR) (Hu & Bentler, 1995).

These goodness-of-fit measures were suggested by Hu and Bentler (1999), that proposed the following cut off values: .95 minimum values for CFI and TLI .08 and .06 maximum values for SRMR and RMSEA, respectively.



SIG= Significance; VAR= Variety; IDE= Identity; AUT= Autonomy; FB= Feedback from the job itself; SFB= Feedback from agents; SAT= Satisfaction; MPS= Motivational Potential Score.

Figure 1. Second-order factor model of teaching adapted version of JDS (Martínez Gómez, Marin-Garcia & Giraldo-O'Meara, 2016)

Then, we employed multigroup confirmatory factor analysis to assess MI between Internet and paper survey modes following the same methodolgy develop by Giraldo-O'Meara et al. (2014). We used the more rigorous, powerful, and versatile multigroup confirmator factor analysis (CFA) approach to assess measurement invariance, which basically determines whether diferent survey settings produce different measures of the same attribute (Steenkamp & Baumgartner, 1998).

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

Consistent with prior research (Vandenberg & Lance, 2002; Vandenberg, 2002), we examined the equality of the observed variance-covariance matrices first.

Satorra-Bentler scaled chi-square adjusted to non-normality ($SB\chi^2$) with robust standard errors (Satorra & Bentler 1994, 2001), the Robust Comparative Fit Index (RCFI), and the Robust Root Mean Square Error of Approximation (RRMSEA) (Curran, West & Finch, 1996; Hu & Bentler, 1999) and Dimitrov (2006), provided the general model fit measurement to assess goodness of fit. Nevertheless, some authors (i.e., Byrne & Stewart 2006; Chen, 2007; Cheung & Rensvold, 2002) argued that it is still possible to use these fit indices to test for measurement equivalence, but focusing on the changes in these measures when adding the constraints at the different steps. They consider that a change larger than .01 is an indication of non-equivalence. We will therefore look at the changes in RRMSEA and RCFI for our different models.

5. Results and discussion

5.1. Variance analysis

Findings achieved to compare web and paper questionnaires, reveal that there is no significant differences across means, except to items s1p04, s1p05, s2p05s and 2p14, which a level of significance lower than 0.05.

5.1. Analysis of scales reliability in each sample

Table 2 shows the values of CR and EV in both samples. They are very close to the recommended value, except for variety that is the feature of teaching methodology, which is measured through items S1P04 and S2P05. The means of these items are different in paper and internet surveys. Values of CR and EV, might be better if we remove this dimension.

Scales	Paper Sample		Web Sample	
	CR[a]	EV[b]	CR	EV
IDE	.5637	.4231	.6127	.4467
VAR	.4873	.3225	.6131	.4461
SIG	.7680	.6254	.87679	.7680
AUT	.7040	.4533	.8030	.6721
FB	.7745	.5440	.8727	.6961
SFB	.7295	.4749	.8593	.6710
SAT	.5761	.5761	.6257	.6257

IDE= Identity; VAR= Variety; SIG= Significance; AUT= Autonomy; SAT= Satisfaction.

Table 2. Construct Reliability (CR) and Extracted Variance (EV) for both models

In addition, we can appreciate that values achieved for both parameters were higher in online surveys, except for the variety factor.

Regarding values of squared correlation coefficients of each item with the relevant factor, responses via web are higher again, except for the relation between the variety dimension and the MPS, as shown in Table 3. These values are appropriate, with values close to .5 or higher, except for items S1P03 and S2P05.

Squared Correlation Coeficients (R^2)	Web	Paper
S1P03	.149	.321
S1P04	.345	.558
S1P05	.512	.648
S1P06	.501	.757
S1P07	.497	.580
S2P04	.520	.758
S2P05	.299	.334
S2P07	.375	.624
S2P09	.563	.571
S2P10	.547	.632
S2P11	.697	.573
S2P12	.584	.750
S2P13	.524	.773
S2P14	.740	.886
S3P03	.423	.626
F1	.638	.330
F2	.406	.856
F3	.413	.492
F4	.480	.554
F5	.683	.655
F6	.559	.754

Table 3. Squared correlation coefficients values

In Table 4, goodness of fit indexes achieved in the web and paper surveys are shown. The values reveal an adequate fit.

Indexes	Web model	Paper model
SB χ^2 [a]	158,0303	126,083
Df [b]	84	84
NFI [c]	.865	.883
NNFI [d]	.913	.945
CFI [e]	.931	.956
IFI [f]	.932	.957
MFI [g]	.880	.865
GFI [h]	.920	.879
AGFI [i]	.885	.827
RMSEA [j]	.064	.073

Table 4. Goodness of fit indexes. SB= chi-square adjusted to non-normality; Df= Degrees of freedom; NFI= Normed Fit Index; NNFI= Not-Nomed Fit Index; CFI= Comparative Fit Index; IFI= Incremental Fit Index; MFI= McDonald Fit Index; GFI= Goodness of Fit Index; AGFI= Adjusted Goodness of fit Index; RMSEA= Root Mean Square Error of Approximation

5.3. Configural invariance

We tested configural invariance across both surveys modes. We began by equality of means, to continue with equality of variances and covariances.

As Table 5 shows, the value of SBχ^2 (p-value = .0000) for means and variances covariances do not support the equaly assumption. In such cases, Satorra and Bentler (1994) proposes to study the invariance of both parameters jointly. This results are shown in Table 6. Although the value of SBχ^2 (p-value = .00282) does not allow establish the hypothesis of invariance, the rest of indexes contradict this conclusion. The Robust Configural Fit Index (RCFI= .982) and the Robust Root Mean Square (RRMSEA= .089) allow us to accept the equality of the number of factors and factor pattern matrices.

Model	χ^2 (p-value)	SB$_\chi{}^2$ (p-value)	Df	RMSEA	CFI	Robust RMSEA	Robust CFI
Equality of means	2653,045 (.00000)	2328,6293 (.00000)	225	.231	n.a	.208	n.a
Equality of covariances and variances	130,070 (**.00000**)	103,7925 (**.00000**)	16	.182	.953	.160	.957

SB= chi-square adjusted to non-normality; Df= Degrees of freedom; CFI= Comparative Fit Index; RMSEA= Root Mean Square Error of Approximation.

Table 5. Goodness of fit indexes for invariance of means and variances-covariances

Model	χ^2 (p-value)	SB$_\chi{}^2$ (p-value)	Df	RMSEA	CFI	Robust RMSEA	Robust CFI
Equality of means and variances	57,228 (.00282)	53,5077 (.00726)	31	.094	.984	.089	.982

SB= chi-square adjusted to non-normality; Df= Degrees of freedom; CFI= Comparative Fit Index; RMSEA= Root Mean Square Error of Approximation.

Table 6. Goodness of fit indexes for invariance of means and variances together

5.4. Metric invariance

As configural invariance is established, we evaluated metric invariance across surveys models, constraining factor loadings in each group. As shown in Tables 7 and 8, the value of SBχ2 change (p-value = .00726) is not significant, but the value of RCFI an RRMSEA allow us to accept that the nested model was still well-fitting. Therefore we could not reject the hypothesis null.

Model	χ^2 (p-value)	SB_χ^2 (p-value)	Df	RMSEA	CFI	Robust RMSEA	Robust CFI
Metric Invariance	258,782 (.00000)	218,2347 (.00005)	143	.061	.952	.049	.963
Metric Invariance without constraints	249,224 (.0000)	205,5941 (.00000)	136	.062	.953	.049	.966

SB= chi-square adjusted to non-normality; Df= Degrees of freedom; CFI= Comparative Fit Index; RMSEA= Root Mean Square Error of Approximation.

Table 7. Goodness of fit indexes for metric invariance

Satorra-Bentler Scaled Difference	D.f.	p-value
16,8065	8	.08136456

Df= Degrees of freedom.

Table 8. Difference of adjusted Satorra-Bentler Chi Squared indexes

5.5. Scalar invariance

Next, we evaluated if scalar invariance can be established constraing the intercepts of both surveys modes. As shown in Tables 9 and 10, the value of $SB\chi^2$ change is very significant (p =5,35907E-11), which indicates that the scalar invariance was not supported. In spite of the change in RCFI and RRMSEA is again lower than 0.1, as the p-value of $SB\chi^2$ change is very close to 0.000, we cannot firmly establish that there are scalar invariance between paper and web survays, but with caution we can accept it.

Model	χ^2 (p-value)	SB_χ^2 (p-value)	Df	RMSEA	CFI	Robust RMSEA	Robust CFI
Scalar Invariance	373,066 (.0000)	289,0748 (.0000)	141	.086	.907	.069	.932
Scalar Invariance without constraints	365,347 (.00000)	208,4460 (.00001)	126	.086	.907	.048	.967

SB= chi-square adjusted to non-normality; Df= Degrees of freedom; CFI= Comparative Fit Index; RMSEA= Root Mean Square Error of Approximation.

Table 9. Goodness of fit indexes for metric invariance

Satorra-Bentler Scaled Difference	D.f.	p-value
80,6288	15	5,35907E-11

Df= Degrees of freedom.

Table 10. Difference of adjusted Satorra-Bentler Chi Squared

5.6. Covariance invariance across latent factors

The next step is to test if there is difference in covariance matrix among latent factors in both groups. Since scalar invariance was not firmly verified, we conducted this test imposing restrictions on the metric invariance model. Results were listed in Table 11. This comparison yielded a value of SBχ^2 change significant, (p-value = .011678944). So, we can state that there is covariance invariance between both groups with a 90% confidence level.

Model	**χ^2 (p-value)**	**SB_χ^2 (p-value)**	**Df**	**RMSEA**	**CFI**	**Robust RMSEA**	**Robust CFI**
Covariance Invariance	279,855 (.00000)	232,9441 (.00001)	147	.059	.942	.047	.954
Metric Invariance	258,782 (.00000)	218,2347 (.00005)	143	.061	.952	.049	.963

SB= chi-square adjusted to non-normality; Df= Degrees of freedom; CFI= Comparative Fit Index; RMSEA= Root Mean Square Error of Approximation.

Table 11. Goodness of fit indexes for covariance invariance

5.7. Variance invariance across latent factors

To evaluate variance invariance across latent factors is neccesary to add a new restricction about the standard errors between both survey modes. If we can establish factor latent variance invariance across groups, as covariance invariance have yet established, correlation across latent factors will be the same in both groups, which will explain that the relation of the factors with the MPS was the same in the original model, independent of the survey mode used. Results are showed in Table 12. As the change of p-value achieved when comparing the SBχ^2 index is .00022376, we cannot firmly establish that variance invariance across latent factors. However, if we observe again the change in the values of RCFI and RRMSEA, we can establish wit caution that latent factors are equivalent in paper and web surveys.

Model	**χ^2 (p-value)**	**SB_χ^2 (p-value)**	**Df**	**RMSEA**	**CFI**	**Robust RMSEA**	**Robust CFI**
Covariance Invariance	279,855 (.00000)	232,9441 (.00001)	147	.059	.942	.047	.954
Latent factors variance invariance	309,505 (.00000)	260,9023 (.00000)	154	.068	.936	.057	.948

SB= chi-square adjusted to non-normality; Df= Degrees of freedom; CFI= Comparative Fit Index; RMSEA= Root Mean Square Error of Approximation.

Table 12. Goodness of fit indexes for latent factors variance invariance

5.8. Variance invariance across errors of latent factors

Finally, we analysed the variance invariance of measurement errors. Results are shown in Table 13. In this case, as the p-value for the change of SBχ^2 is .02190079, we can state that reliability of the surveys items is similar between online and paper-based surveys.

Model	χ^2 (p-value)	SB$_\chi$^{2} (p-value)	Df	RMSEA	CFI	Robust RMSEA	Robust CFI
Invariance factors variance	309,505 (.00000)	260,9023 (.00000)	154	.068	.936	.057	.948
Invariance errors variance	355,127 (.00000)	287,4699 (.00000)	168	.072	.927	.057	.935

SB= chi-square adjusted to non-normality; Df= Degrees of freedom; CFI= Comparative Fit Index; RMSEA= Root Mean Square Error of Approximation.

Table 13. Goodness of fit indexes for errors variance invariance of latent factors

The research results confirm the equivalence in respect to the paper and web surveys of JDS adapted to university teaching, reveal the same factor structure, factor loadings and reliability scales. Bartram (2005) already argued this requirement when he stated that if a research is collecting data from distinct survey modes is necessary to test equivalence across them. In addition, it is a recommendation of the Standards for Educational Research Association (American Psychological Association and National Council on Measurement in Education, 1999). In our case, we have not had scalar invariance, but it is only a requirement when comparing means of latent factors, because it would mean that comparisons of this parameter across groups could be biased due to differences on scales and data sources (Cheung & Rensvold, 2002). Besides, we have not achieved a significant change of SBχ^2 value for variance invariance across factors latent, the rest of indexes contradict that conclusion and we can accept with caution that there is complete invariance in both contexts. These results confirm identical psychometric properties for online and paper modes of JDS, according with findings in other studies with different surveys (i.e. Drasgow & Schmidt, 2002; Martins, 2010; Meade et al., 2007).

6. Conclusions, limitations and future research

The main purpose of our study was to evaluate if the underlying factor structure of the teaching version of JDS was equivalent with data collected from on line and paper surveys. According the results, there is no differences between data collected with web and paper based surveys. The style of collecting date did not seem to have an influence in terms of the construct measures. Metric invariance, covariance invariance, variance of latent factors invariance and measurement errors invariance can be established between two groups. The non-fulfilment of scalar invariance only affects when comparing means across factors, but if the research target is to see if there are significant relations across variables, scalar variance is not important. As Van de Schoot, Lugtig and Hox (2012) set out, when checking if factor loadings, items coefficients and residual variances are equivalent across groups, we can state that comparisons made across groups are valid at all levels.

These findings have practical implications as well, since they add a new sample in which web and paper questionnaires are equivalent and for teachers to desire to change the teaching methodology at university, encourage students' participation and teamwork through active methodologies. The cultural context of students (different degrees and academic years) has been testing before (Martínez Gómez et al., 2016), according the recommendations of Byrne and Van De Vijver (2010, p.128), where they state that "testing for equivalence of a measuring instrument in cross-cultural studies can be fraught with difficulties".

There were of course, limitations in this study. As stated previously, we used a student sample with a specific questionnaire and the generalisation to other questionnaire, or population, should be proved with specific data and we should cautiously interpret the equivalence of web and paper-based surveys. The generalisation of our findings to other context would allow to aggregate information obtained of different types of surveys, increasing the number of responses obtained in researches.

We should note that our sample size is rather small for SEM models with such numbers of estimated parameters. According Kline (2010) the typical sample size of 200 cases in studies where SEM is too small when analyzing a complex model using an estimation method other the ML or distributions non-normal. It is possible that if analyses samples larger sizes or different context or universities, we would yield different results.

Based on the present results study, we recommended that the target of the future studies should be tested the invariance in first-order models following the same sequence. Secondly, the lines should be drawn for all universities extend the sample to a representative population of the university students (Spanish or other countries). In this case, the instrument validity should increase.

Acknowledgements

This paper has been written with financial support from the Project "Validación de las Competencias Transversales de Innovación mediante un enfoque formativo". (GV/2016/004) de la Conselleria d'Educació, Investigació, Cultura i Esport (Generalitat Valenciana).

References

American Psychological Association & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Aster, A.Z. (2004). Consumer research goes online. *Marketing Magazine*, 7, 13-14.

Aydin, B., & Ceylan, A. (2008). The employee satisfaction in metalworking manufacturing: How do organizational culture and organizational learning capacity jointly affect it?. *Journal of Industrial Engineering and Management*, 1(2), 143-168.

Barak, M., Ben-Chaim D., & Zoller, U. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education*, 37, 353-369. https://doi.org/10.1007/s11165-006-9029-2

Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, 90, 1185-1203. https://doi.org/10.1037/0021-9010.90.6.1185

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606. https://doi.org/10.1037/0033-2909.88.3.588

Bollen, K.A., & Long, J.S. (1993). *Testing Structural Equation Models*. Newbury Park, California: Sage.

Bosnjak, M., Tuten, T.L., & Wittmann, W.W. (2005). Unit (non) response inweb-based access panel surveys: an extended planned-behavior approach. *Psychology & Marketing* 22(6), 489-505. https://doi.org/10.1002/mar.20070

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, 27, 281-291. https://doi.org/10.1093/pubmed/fdi031

Brown, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp.136-162). Newbury Park, California: Sage.

Buchanan, T., & Smith, J.L. (1999). Using the internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, 90, 125-144. https://doi.org/10.1348/000712699161189

Byrne, B.M. (1989). Multigroup comparisons and the assumption of equivalent construct validity across groups: Methodological and substantive issues. Multivariate Behavioural Research, 24, 503-523. https://doi.org/10.1207/s15327906mbr2404_7

Byrne, B.M., & Stewart, S.M. (2006).The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, 13, 287-321. https://doi.org/10.1207/s15328007sem1302_7

Byrne, B.M., & Van De Vijver, F.J.R. (2010). Testing for measurement and structural equivalence in large-scalecross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10, 107-132. https://doi.org/10.1080/15305051003637306

Chen, F.F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling*, 14, 464-504. https://doi.org/10.1080/10705510701301834

Chen, F.F.; Sousa, K.H.; West, S.G. (2005). Testing Measurement Invariance of Second-Order Factor Models. *Structural Equation Modeling*, 12: 471-492. https://doi.org/10.1207/s15328007sem1203_7

Cheung, G.W. (2008). Testing equivalence in the structure, means, and variances of higher-order constructs with structural equation modelin. *Organizational Research Methods*, 11(3), 593-613. https://doi.org/10.1177/1094428106298973

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indices for testing measurement Equivalence. *Structural Equation Modeling*, 9, 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385-396. https://doi.org/10.2307/2136404

Cole, M.S., Bedeian, A.G., & Feild, H.S. (2006). The measurement equivalence of web-based and paper-and-pencil measures of transformation leadership: A multinational test. *Organisational Research Methods*, 9(2), 339-368. https://doi.org/10.1177/1094428106287434

Cook, C., Heath, F., & Thompson, R.L. (2000). A meta-analysis of response rates in Web- or Internet based surveys. *Educational and Psychological Measurement*, 60, 821-836. https://doi.org/10.1177/00131640021970934

Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(11), 16-29. https://doi.org/10.1037/1082-989X.1.1.16

Davidov, E., & Depner, F. (2011). Testing for measurement equivalence of human values across online and paper-and pencil surveys. *Quality & Quantity*, 45(2), 375-390. https://doi.org/10.1007/s11135-009-9297-9

De Beuckelaer, A., & Lievens, F. (2009). Measurement equivalence of paper‐and‐pencil and Internet organisational surveys: A large scale examination in 16 countries. *Applied Psychology*, 58(2), 336-361. https://doi.org/10.1111/j.1464-0597.2008.00350.x

Dillman, D.A. (2000). *Mail and Internet Surveys: The Tailored Design Method* (2nd Eds.). New York: John Wiley & Sons.

Dillman, D., Smyth, J., & Christian, L. (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New York: Wiley.

Dimitrov, D.M. (2006). Validation of cognitive operations and processes across ability levels and individual test items. In T.E. Scruggs & M.A. Mastropieri, (Eds.), *Advances in Learning and Behavioral Disabilities* (pp. 55-81). San Diego, CA: Elsevier. Ltd. https://doi.org/10.1016/S0735-004X(06)19003-5

Drasgow, F., & Schmidt, N. (2002). *Measuring and Analyzing Behaviour in Organizations: Advances in Measurement and Data Analysis*. San Francisco: Jossey Bass.

Ebenezer, J.V., Columbus, R., Kaya, O.N., Zhang, L., & Ebenezer, D.L. (2012). One science teacher's professional development experience: A case study exploring changes in students' perceptions of their fluency with innovative tecnologies. *Journal of Science Educatonal Technologies,* nº, xx-xx. Retrieved from: http://www.springerlink.com/content/q03j2118040t6863

Elosua, P. (2005). Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de autoconcepte. *Psicothema*, 17(2), 356-362.

Epstein, J., Klinkenberg, W.D., Wiley, D., & McKinley, L. (2001). Ensuring sampling equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior*, 17, 339-346. https://doi.org/10.1016/S0747-5632(01)00002-4

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26(2), 132-139. https://doi.org/10.1016/j.chb.2009.10.015

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. https://doi.org/10.1089/cyber.2011.0411

Fang, J., Wen, C., & Prybutok, V.R. (2014). An assessment of equivalence between internet and paper-based surveys: Evidence from collectivistic cultures. *Quality & Quantity*, 48(1), 493-506. https://doi.org/10.1007/s11135-012-9783-3

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. https://doi.org/10.1177/10791102009002011

Gangestad, S., & Snyder, M. (1985). To carve nature at its joints': On the existence of discrete classes in personality. *Psychological Review*, 92, 317-349. https://doi.org/10.1037/0033-295X.92.3.317

Giraldo-O'Meara, M., Marin-Garcia, J.A., & Martínez-Gómez, M. (2014). Validation of the JDS satisfaction scales applied to educational university environments. *Journal of Industrial Engineering and Management*, 7(1), 72-99. https://doi.org/10.3926/jiem.906

Göritz, A.S. (2006). Incentives in web studies: methodological issues and a review. *International Journal of Internet Science*, 1(1), 58-70.

Hackman, J.R., & Oldham, G.R. (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology*, 60(2), 159-170. https://doi.org/10.1037/h0076546

Hackman, J.R., & Oldham, G.R. (1976). Motivation through the design of the work: Test of a theory. *Organizational Behaviour and Human Performance*, 16, 250-279. https://doi.org/10.1016/0030-5073(76)90016-7

Hackman, J.R., & Oldham, G.R. (1980). *Work Redesig*. Reading, MA: Addison-Wesley.

Hair, J.F., Anderson, R.E., Thatam, R.L., & Black, W.C. (1998). *Multivariate Data Analysis* (6th eds). New York: Prentice Hall International.

Herrero, J., & Meneses, J. (2006). Short Web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CESD) Scales: A comparison to pencil and paper responses among Internet users. *Computers in Human Behavior*, 22(5), 830-846. https://doi.org/10.1016/j.chb.2004.03.007

Hogg, A. (2003). Web efforts energize customer research. *Electronic Perspectives*, 6, 81-83.

Hohwü, L., Lyshol, H., Gissler, M., Jonsson, S.H., Petzold, M., & Obel, C. (2013). Web-based versus traditional paper questionnaires: A mixed-mode survey with a Nordic perspective. *Journal of Medical Internet Research*, 15(8), e173. https://doi.org/10.2196/jmir.2595

Hu, L., & Bentler, P.M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural Equation Modeling. Concepts, Issues, and Applications* (pp.76-99). London: Sage.

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. https://doi.org/10.1080/10705519909540118

Ismail, A., Mashkuri, A., Sulaiman, A., & Kee Hock, W. (2011). Interactional justice as a mediator of the relationship between pay for performance and job satisfaction. *Intangible Capital*, 7(2), 213-235.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426. https://doi.org/10.1007/BF02291366

King, W., & Miles, E. (1995). Quasi-experimental assessment of the effect of computerizing noncognitive paper and pencil measurements: A test of measurement equivalence. *Journal Applied of Psycholy*, 80(6), 643-651. https://doi.org/10.1037/0021-9010.80.6.643

Kline, R.B. (2010). *Principles and Practice of Structural Equation Modeling*. NY, London: The Guilford Press.

Kraut, A.I., & Saari, L.M. (1999). Organizational surveys: Coming of age for a new era. In A.I. Kraut & A.K. Korman (Eds), *Evolving practices in human resource management* (pp. 302-327). San Francisco, CA: Jossey-Bass.

Lautenschlager, G.J., & Flaherty, V.L. (1990). Computer administration of questions: More desirable or more social desirability?. *Journal of Applied Psychology*, 75, 310-314. https://doi.org/10.1037/0021-9010.75.3.310

Leung, D., & Kember, D. (2005). Comparability of data gathered from evaluation questionnaires on paper and through the internet. *Research in Higher Education*, 46(5), 571-591. https://doi.org/10.1007/s11162-005-3365-3

Marbach-Ad, G., & Sokolove, P.G. (2002). The use of e-mail and in-class writing to facilitate student–instructor interaction in large-enrollment traditional and active learning classes. *Journal of Science Education Technology*, 11 (2), 109-119. https://doi.org/10.1023/A:1014609328479

Martínez-Gomez, M., & Marin-Garcia, J.A. (2009). Como medir y guiar el cambio hacia entornos educativos universitarios más motivadores para los alumnos. *Formación Universitaria*, 2, 3-14. https://doi.org/10.4067/S0718-50062009000400002

Martínez Gómez, M., Marin-Garcia, J., & Giraldo-O'Meara, M. (2016). The measurement invariance of job diagnostic survey (jds) across three university student groups. *Journal of Industrial Engineering and Management*, 9(1), 17-34. https://doi.org/10.3926/jiem.1783

Martins, N. (2010). Measurement model equivalence in web-and paper-based surveys. *Sourthen African Business Review*, 14(3), 77-107.

Meade, A.W., Michels, L.C., & Lautenschlager, G.J. (2007). Are internet and paper-and pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods*, 10(2), 322-345. https://doi.org/10.1177/1094428106289393

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, 66-74. https://doi.org/10.1177/0013164498058001006

Nulty, D.D. (2008). The adequacy of response rates to online and paper surveys: What can be done?. *Assessment & Evaluation in Higher Education*, 33(3), 301-314. https://doi.org/10.1080/02602930701293231

Orgambídez-Ramos, A., Borrego-Alés, Y., & Mendoza-Sierra, I. (2014). Role stress and work engagement as antecedents of job satisfaction in Spanish workers. *Journal of Industrial Engineering and Management*, 7(1), 360-372. https://doi.org/10.3926/jiem.992

Radloff, L. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401. https://doi.org/10.1177/014662167700100306

Riva, G., Teruzzi, T., & Anolli, L. (2003). The use of the internet in psychological research: Comparison of online and offline questionnaires. *Cyber Psychology & Behavio*r, 6, 73-80. https://doi.org/10.1089/109493103321167983

Reips, U.D. (2000). The web experiment method: Advantages, disadvantages and solutions. In M.H. Birnbaum (Eds.), *Psychological Experiments on the Internet* (pp. 89-117). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-012099980-4/50005-8

Roberts, L.L., Konczak, L.J., & Macan, T.H. (2004). Effects of data collection method on organizational climate survey results. *Applied H.R.M Research*, 9, 13-26.

Santos Rego, M.Á., Godás Otero, A., Lorenzo Moledo, M., & Gómez Fraguela, J.A. (2010). Eficacia y satisfacción laboral de dos profesores no universitarios: Revisión de un instrumento de medida. *Revista Española de Pedagogía*, 245, 151-168.

Satorra, A., & Bentler, P.M. (1994). Corrections to test statistic and standard errors in covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), *Analysis of Latent Variables in Developmental Research* (pp. 399-419). Thousand Oaks, CA: Sage.

Satorra, A., & Bentler, P.M. (2001). A Scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514. https://doi.org/10.1007/BF02296192

Schaeffer, D., & Dillman, D. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly*, 62, 378-397. https://doi.org/10.1086/297851

Schonlau, M., Fricker, R.D., & Elliott, M.N. (2002). *Conducting research surveys via e-mail and the web*. Santa Monica, CA: Rand Corporation.

Simsek, Z., & Veigha, J.F. (2001). A primer on internet organizational surveys. *Organizational Research Methods*, 4, 218-235. https://doi.org/10.1177/109442810143003

Sproull, L.S. (1986). Using electronic mail for data collection in organizational research. *Academy of Management Journal*, 29, 159-169. https://doi.org/10.2307/255867

Stanton, J.M. (1998). An empirical assessment of data collection using the internet. *Personnel Psychology*, 51, 709-725. https://doi.org/10.1111/j.1744-6570.1998.tb00259.x

Steenkamp, J.B.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in crossnational consumer research. *Journal of Consumer Research*, 25, 78-90. https://doi.org/10.1086/209528

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. (2009). Testing measurement invariance using multigroup CFA: differences between educational groups in human values measurement. *Quality & Quantity*, 43(4), 599-616. https://doi.org/10.1007/s11135-007-9143-x

Trullas, I., & Enache, M. (2011). Theoretical analysis of the antecedents and the consequences of students' identification with their university and their perception of quality. *Intangible Capital*, 7(1), 170-212. https://doi.org/10.3926/ic.2011.v7n1.p170-212

Ullman, J.B., & Bentler, P.M. (2004) Structural Equation Modeling. In M. Hardy & Bryman (Eds.), *Handbook of Data Analysis* (pp.431-458). London: Sage.

Van Gelder, M.M., Bretveld R.W., & Roeleveld, N. (2010). Web-based questionnaires: The future in epidemiology?. *American Journal of Epidemiol*, 172(11), 1292-1298. https://doi.org/10.1093/aje/kwq291

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

Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis on the measurement invariance literature: Suggestions, practices and recommendations for organisational research. *Organizational Research Methods*, 3, 4-70. https://doi.org/10.1177/109442810031002

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486-492. https://doi.org/10.1080/17405629.2012.686740

Walt, N., Atwood, K., & Mann, A. (2008). Does Survey Medium Affect Responses? An Exploration of Electronic and Paper Surveying in British Columbia Schools. *Journal of Technology, Learning, and Assessment*, 6(7). Retrieved from: http://www.jtla.org/

Young, S.A., Daum, D.L., Robie, C., & Macey, W.H. (2000). Paper vs. web survey administration: Do different methods yield different results?. *Proccedings of 15th Anual Conference of the Society for Industrial and Organizational Psychology*, New Orleans, LA.

Yu, S.C., & Yu, M.N. (2007). Comparison of Internet-based and paper-based questionnaires in Taiwan using multisample invariance approach. *Cyberpsycholy Behaviour*, 10(4), 501-507. https://doi.org/10.1089/cpb.2007.9998

Yun, G.W. & Trumbo, C.W. (2000). Comparative Response to a Survey Executed by Post, e-mail, & Web Form. *Journal of Computer-Mediated Communication*, 6(1). Retrieved from: http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2000.tb00112.x/full https://doi.org/10.1111/j.1083-6101.2000.tb00112.x