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

Monica Martinez-Gomez1, Juan A. Marín-García2, Martha Giraldo O’Meara3

1Departamento de Estadística e Investigación Operativa. Universidad Politécnica de Valencia (Spain)
2ROGLE. Departamento de Organización de Empresas. Universitat Politècnica de València (Spain)
3Universitat de València (Spain)

momargo@eio.upv.es, jamarin@omp.upv.es, marthasmeara@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 assess 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 prerequisite 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, Breteveld 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

- 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.g., 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.g., Dillman Smyth & Christian, 2009; Martins, 2010).
- Less costly (e.g., Dillman, 2000; Kraut & Saari, 1999; Schaeffer & Dillman, 1998; Sproull, 1986; Young, Daum, Robie & Macey, 2000; Yun & Trumbo, 2000).
- Lead to faster survey responses (e.g., Schaeffer & Dillman, 1998; Sproull, 1986).
- Allow greater flexibility in survey design (Dillman, 2000).
- Offer a wider variety of response formats (Simsek & Veiga, 2001).
- There is no human (coding) errors (e.g., Cook et al., 2000; Roberts, Konczak & Macan, 2004).
- Are free of experimenter bias (e.g., 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).

Disadvantages of web surveys

- 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.g., Dillman et al., 2009; Martins, 2010) (ID Systems= Autoidentification Systems).
- Higher non-response rates e.g., Schaeffer & Dillman, 1998; Sproull, 1986).
- Potential technological problems (Kraut & Saari, 1999).
- Decreased item reliability due to somewhat higher measurement errors (Stanton, 1998).
- Possibility of multiple submissions (Reips, 2000).

<table>
<thead>
<tr>
<th>Advantages of web surveys</th>
<th>Disadvantages of web surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.g., Schonlau, Fricker &amp; Elliott, 2002).</td>
<td>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.g., Dillman et al., 2009; Martins, 2010) (ID Systems= Autoidentification Systems).</td>
</tr>
<tr>
<td>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.g., Dillman Smyth &amp; Christian, 2009; Martins, 2010).</td>
<td>Higher non-response rates e.g., Schaeffer &amp; Dillman, 1998; Sproull, 1986).</td>
</tr>
<tr>
<td>Less costly (e.g., Dillman, 2000; Kraut &amp; Saari, 1999; Schaeffer &amp; Dillman, 1998; Sproull, 1986; Young, Daum, Robie &amp; Macey, 2000; Yun &amp; Trumbo, 2000).</td>
<td>Higher probability of getting dishonest answers (Lautenschlager &amp; Flaherty, 1990).</td>
</tr>
<tr>
<td>Lead to faster survey responses (e.g., Schaeffer &amp; Dillman, 1998; Sproull, 1986).</td>
<td>Potential technological problems (Kraut &amp; Saari, 1999).</td>
</tr>
<tr>
<td>Allow greater flexibility in survey design (Dillman, 2000).</td>
<td>Decreased item reliability due to somewhat higher measurement errors (Stanton, 1998).</td>
</tr>
<tr>
<td>Offer a wider variety of response formats (Simsek &amp; Veiga, 2001).</td>
<td>Possibility of multiple submissions (Reips, 2000).</td>
</tr>
<tr>
<td>Wider geographical reach (Epstein, Klinkenberg, Wiley &amp; McKinley, 2001).</td>
<td></td>
</tr>
<tr>
<td>There is no human (coding) errors (e.g., Cook et al., 2000; Roberts, Konczak &amp; Macan, 2004).</td>
<td></td>
</tr>
<tr>
<td>Are free of experimenter bias (e.g., Reips, 2000).</td>
<td></td>
</tr>
<tr>
<td>Are less sensitive to order of question effects due to the ease of randomising questions (Bowling, 2005).</td>
<td></td>
</tr>
<tr>
<td>Not have many missing values (Stanton, 1998).</td>
<td></td>
</tr>
<tr>
<td>Greater flexibility in survey design and wider variety of response formats (Simsek &amp; Veiga, 2001).</td>
<td></td>
</tr>
<tr>
<td>Wider geographical reach (Epstein et al., 2001).</td>
<td></td>
</tr>
</tbody>
</table>

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. Initially all of these studies used traditional techniques to analyse the equivalence between different groups. For example, 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), achieving 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 parameters 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 Depner (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; Orgambidez-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 students 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, Marín-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 stimulating a job position. As we detailed 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 potential 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 Fraguera, 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
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 JD (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 methodology 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 different 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.

<table>
<thead>
<tr>
<th>Scales</th>
<th>Paper Sample</th>
<th>Web Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CR$^a$</td>
<td>EV$^b$</td>
</tr>
<tr>
<td>IDE</td>
<td>.5637</td>
<td>.4231</td>
</tr>
<tr>
<td>VAR</td>
<td>.4873</td>
<td>.3225</td>
</tr>
<tr>
<td>SIG</td>
<td>.7680</td>
<td>.6254</td>
</tr>
<tr>
<td>AUT</td>
<td>.7040</td>
<td>.4533</td>
</tr>
<tr>
<td>FB</td>
<td>.7745</td>
<td>.5440</td>
</tr>
<tr>
<td>SFB</td>
<td>.7295</td>
<td>.4749</td>
</tr>
<tr>
<td>SAT</td>
<td>.5761</td>
<td>.5761</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Squared Correlation Coefficients (R²)</th>
<th>Web</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1P03</td>
<td>.149</td>
<td>.321</td>
</tr>
<tr>
<td>S1P04</td>
<td>.345</td>
<td>.588</td>
</tr>
<tr>
<td>S1P05</td>
<td>.512</td>
<td>.648</td>
</tr>
<tr>
<td>S1P06</td>
<td>.501</td>
<td>.757</td>
</tr>
<tr>
<td>S1P07</td>
<td>.497</td>
<td>.580</td>
</tr>
<tr>
<td>S2P04</td>
<td>.520</td>
<td>.758</td>
</tr>
<tr>
<td>S2P05</td>
<td>.299</td>
<td>.334</td>
</tr>
<tr>
<td>S2P07</td>
<td>.375</td>
<td>.624</td>
</tr>
<tr>
<td>S2P09</td>
<td>.563</td>
<td>.571</td>
</tr>
<tr>
<td>S2P10</td>
<td>.547</td>
<td>.632</td>
</tr>
<tr>
<td>S2P11</td>
<td>.697</td>
<td>.573</td>
</tr>
<tr>
<td>S2P12</td>
<td>.584</td>
<td>.750</td>
</tr>
<tr>
<td>S2P13</td>
<td>.324</td>
<td>.773</td>
</tr>
<tr>
<td>S2P14</td>
<td>.740</td>
<td>.886</td>
</tr>
<tr>
<td>S3P03</td>
<td>.423</td>
<td>.626</td>
</tr>
<tr>
<td>F1</td>
<td>.638</td>
<td>.330</td>
</tr>
<tr>
<td>F2</td>
<td>.406</td>
<td>.856</td>
</tr>
<tr>
<td>F3</td>
<td>.413</td>
<td>.492</td>
</tr>
<tr>
<td>F4</td>
<td>.480</td>
<td>.554</td>
</tr>
<tr>
<td>F5</td>
<td>.683</td>
<td>.655</td>
</tr>
<tr>
<td>F6</td>
<td>.539</td>
<td>.754</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Indexes</th>
<th>Web model</th>
<th>Paper model</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB χ² a</td>
<td>158,0303</td>
<td>126,083</td>
</tr>
<tr>
<td>Df b</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>NFI c</td>
<td>.865</td>
<td>.883</td>
</tr>
<tr>
<td>NNFI d</td>
<td>.913</td>
<td>.945</td>
</tr>
<tr>
<td>CFI e</td>
<td>.931</td>
<td>.956</td>
</tr>
<tr>
<td>IFI f</td>
<td>.932</td>
<td>.957</td>
</tr>
<tr>
<td>MFI g</td>
<td>.880</td>
<td>.865</td>
</tr>
<tr>
<td>GFI h</td>
<td>.920</td>
<td>.879</td>
</tr>
<tr>
<td>AGFI i</td>
<td>.885</td>
<td>.827</td>
</tr>
<tr>
<td>RMSEA j</td>
<td>.064</td>
<td>.073</td>
</tr>
</tbody>
</table>

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 $SB\chi^2$ (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 $SB\chi^2$ (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.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$ (p-value)</th>
<th>$SB\chi^2$ (p-value)</th>
<th>Df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>Robust RMSEA</th>
<th>Robust CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equality of means</td>
<td>2653,045 (.00000)</td>
<td>2328,6293 (.00000)</td>
<td>225</td>
<td>.231</td>
<td>n.a</td>
<td>.208</td>
<td>n.a</td>
</tr>
<tr>
<td>Equality of covariances and variances</td>
<td>130,070 (.00000)</td>
<td>103,7925 (.00000)</td>
<td>16</td>
<td>.182</td>
<td>.953</td>
<td>.160</td>
<td>.957</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$ (p-value)</th>
<th>$SB\chi^2$ (p-value)</th>
<th>Df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>Robust RMSEA</th>
<th>Robust CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equality of means and variances</td>
<td>57,228 (.00282)</td>
<td>53,5077 (.00726)</td>
<td>31</td>
<td>.094</td>
<td>.984</td>
<td>.089</td>
<td>.982</td>
</tr>
</tbody>
</table>

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 $SB\chi^2$ 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.
5.5. Scalar invariance

Next, we evaluated if scalar invariance can be established constraining 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 surveys, but with caution we can accept it.
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 SBχ² change significant, (p-value = .011678944). So, we can state that there is covariance invariance between both groups with a 90% confidence level.

<table>
<thead>
<tr>
<th>Model</th>
<th>χ² (p-value)</th>
<th>SBχ² (p-value)</th>
<th>Df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>Robust RMSEA</th>
<th>Robust CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance Invariance</td>
<td>279,855 (.00000)</td>
<td>232,9441 (.00001)</td>
<td>147</td>
<td>.059</td>
<td>.942</td>
<td>.047</td>
<td>.954</td>
</tr>
<tr>
<td>Metric Invariance</td>
<td>258,782 (.00000)</td>
<td>218,2347 (.00005)</td>
<td>143</td>
<td>.061</td>
<td>.952</td>
<td>.049</td>
<td>.963</td>
</tr>
</tbody>
</table>

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 necessary to add a new restriction 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 SBχ² 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 with caution that latent factors are equivalent in paper and web surveys.

<table>
<thead>
<tr>
<th>Model</th>
<th>χ² (p-value)</th>
<th>SBχ² (p-value)</th>
<th>Df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>Robust RMSEA</th>
<th>Robust CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance Invariance</td>
<td>279,855 (.00000)</td>
<td>232,9441 (.00001)</td>
<td>147</td>
<td>.059</td>
<td>.942</td>
<td>.047</td>
<td>.954</td>
</tr>
<tr>
<td>Latent factors variance invariance</td>
<td>309,505 (.00000)</td>
<td>260,9023 (.00000)</td>
<td>154</td>
<td>.068</td>
<td>.936</td>
<td>.057</td>
<td>.948</td>
</tr>
</tbody>
</table>

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 $SB\chi^2$ is .02190079, we can state that reliability of the surveys items is similar between online and paper-based surveys.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$ (p-value)</th>
<th>$SB\chi^2$ (p-value)</th>
<th>Df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>Robust RMSEA</th>
<th>Robust CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invariance factors</td>
<td>309,505 (.00000)</td>
<td>260,9023 (.00000)</td>
<td>154</td>
<td>.068</td>
<td>.936</td>
<td>.057</td>
<td>.948</td>
</tr>
<tr>
<td>Invariance errors</td>
<td>355,127 (.00000)</td>
<td>287,4699 (.00000)</td>
<td>168</td>
<td>.072</td>
<td>.927</td>
<td>.057</td>
<td>.935</td>
</tr>
</tbody>
</table>

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 $SB\chi^2$ 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


-896-


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


---

Article's contents are provided on an Attribution-Non Commercial 3.0 Creative commons license. Readers are allowed to copy, distribute and communicate article's contents, provided the author's and Intangible Capital's names are included. It must not be used for commercial purposes. To see the complete license contents, please visit http://creativecommons.org/licenses/by-nc/3.0/.

-901-