Measurement Invariance of the Pay Satisfaction Questionnaire Across Three Countries

Filip Lievens
Frederik Anseel
Ghent University
Michael M. Harris
University of Missouri–St. Louis
Jacob Eisenberg
University College Dublin

In recent years, pay satisfaction has been increasingly studied in an international context, prompting the importance of examining whether the Pay Satisfaction Questionnaire (PSQ) is invariant across countries other than the United States. This study investigated the measurement invariance across three countries, namely, the United States (N = 321), Belgium (N = 301), and Cyprus (N = 132). Results showed that the measurement structure of the PSQ was invariant across these different countries because there was no departure from measurement invariance in terms of factor form, factor pattern coefficients, factor variances, and factor covariances. These results show promise for the equivalence of PSQ ratings across different countries. Future research is needed to test the equivalence further across other countries and samples.

Keywords: pay satisfaction; measurement invariance; confirmatory factor analysis; culture

The construct of pay satisfaction, as well as its determinants, has been the focus of much research during the last few decades. One of the most important developments in this area in the last 20 years was the development of the Pay Satisfaction Questionnaire (PSQ; H. G. Heneman & Schwab, 1985). The development of the PSQ and subsequent research using the PSQ led to the realization that pay satisfaction was a multidimensional construct (e.g., Scarpello, Huber, & Vandenberg, 1988), with different factors having different antecedents and different outcomes (e.g., Judge, 1993). Generally, the expected factor structure of the PSQ (i.e., pay level, pay raise, pay administration, and benefits) has been supported across various studies in the United

Authors’ Note: The authors thank Stijn Veldeman, Ines Vandamme, Daina Nicolaou, and Nicos Constantinou for their help in collecting the data. Please address correspondence to Filip Lievens, Department of Personnel Management, Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium; e-mail: filip.lievens@ugent.be.
States (e.g., DeConinck, Stilwell, & Brock, 1996; R. L. Heneman, Greenberger, & Strasser, 1988; Judge, 1993; Judge & Welbourne, 1994).

Despite the relatively large literature on the dimensionality of pay satisfaction, H. G. Heneman and Judge (2000) pointed out a number of shortcomings in this area. Specifically, they noted that “almost all of the studies on pay satisfaction dimensionality have been conducted on American samples” (p. 84); therefore, it would be of value to examine whether similar results are found in other cultures and countries. The need to examine the PSQ in an international context is further highlighted by the increasing number of studies that have examined pay satisfaction across countries (Fong & Schaffer, 2003; Sweeney & McFarlin, 2004). In a similar vein, there is an increasing trend of conducting pay level and satisfaction surveys across countries due to the globalization of the economy.

A problem is that many of these studies and surveys use the PSQ across countries without investigating whether the measurement structure underlying the PSQ scores is invariant. However, when the PSQ is used in an international context, it is not justified to compare PSQ ratings across countries without establishing that the measurement structure underlying the PSQ scores is equivalent across these countries. Along these lines, Hoyle and Smith (1994) concluded that a comparison between groups on the basis of observed mean differences alone on scores whose reliability and factorial validity have not been proven to be (at least partially) invariant across these groups can be misleading and is a classic example of “comparing apples and oranges” (p. 433).

The purpose of this study is to examine the factor structure of the PSQ across different countries and to establish measurement invariance. Specifically, we used multiple-group confirmatory factor analysis (CFA) to conduct a sequence of increasingly more restrictive tests of invariance across three countries. An examination of measurement invariance enables one to determine whether the items and the underlying constructs of the PSQ mean the same thing to respondents of these different countries (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000).

The PSQ

In early studies, pay satisfaction was considered to be a unidimensional construct. Whereas some researchers developed ad hoc measures for measuring pay satisfaction, other researchers relied on more standardized instruments for specifically targeting pay level satisfaction such as the pay satisfaction subscales of the Minnesota Satisfaction Questionnaire and the Job Descriptive Index.

A major breakthrough in the study of pay satisfaction was H. G. Heneman and Schwab’s (1985) conceptualization of pay satisfaction as a multidimensional construct and the development of the PSQ. Initially, H. G. Heneman and Schwab suggested that pay satisfaction exists along five relatively independent dimensions: pay level, pay administration, pay structure, pay raise, and benefits. However, confirmatory factor
analytic results revealed that the pay structure and pay administration dimensions could not accurately be distinguished from each other. Additional exploratory factor analyses suggested that a four-factor solution was more appropriate. Therefore, the original PSQ was modified to an 18-item measure that tapped four dimensions (pay level, benefits, pay raise, and pay structure/administration).

Over the years, various studies have examined the factor structure underlying the PSQ (Brown & Huber, 1992; Carraher, 1991; Carraher & Buckley, 1996; DeConinck et al., 1996; Fong & Shaffer, 2003; R. L. Heneman et al., 1988; Judge, 1993; Judge & Welbourne, 1994; Lam, 1998; Mulvey, Miceli, & Near, 1992; Orpen & Bonnici, 1987; Scarpello et al., 1988). As summarized by H. G. Heneman and Judge (2000), these factor analytic studies have supported the notion of pay satisfaction as a multidimensional construct. The multidimensionality of pay satisfaction is further supported by evidence of differential prediction (Judge, 1993).

Although most studies using confirmatory analysis found support for the expected four-factor structure (e.g., DeConinck et al., 1996; Judge, 1993; Judge & Welbourne, 1994), not all studies reported consistent results concerning the structure and administration items, leading to questions about what these items are actually measuring. One explanation for some of the inconsistent results obtained is that the factor structure, or number of factors, of the PSQ might depend on contextual characteristics. For instance, Scarpello et al. (1988) found that the PSQ factor structure varied by job classification (exempt, nonexempt, and hourly) and by other organizational factors such as company labor-relations policy.

Thus, the PSQ has been studied in a wide variety of organizations, with many different employee samples. For the most part, the empirical evidence shows that four factors are the best representation of PSQ ratings in the United States. The finding that the factor structure of the PSQ might vary according to contextual factors further shows the need for assessing measurement equivalence across different cultural contexts.

Therefore, we gathered PSQ data from three different countries (i.e., the United States, Belgium, and Cyprus) to examine the measurement invariance of the PSQ in an international context. The choice of these countries was guided by theoretical considerations. More specifically, the countries in this study were included because they belonged to three different cultural regions. According to Schwartz’s (1999, pp. 36, 39) categorization, the United States belongs to the English-speaking cultural region, Belgium to the West European cultural region, and Cyprus to the East European cultural region.

Method

Participants and Procedure

To ensure comparability across our three countries, we made certain that all respondents were employed and had prior work experience. This work experience
provided them with the background necessary for rating their pay satisfaction on the basis of the PSQ.

**U.S. sample.** These participants were recruited from a variety of organizations in the United States. To this end, we used a survey panel (see www.studyresponse.com for details about this system) that has a cross-sample of respondents from a variety of organizations. There were 321 respondents in the U.S. sample (56% females and 44% males), which reflects approximately a 30% response rate. The mean age of the respondents was 36.3 years. Their mean work experience was 15.2 years. The two most dominant job categories were clerical jobs (33%) and technical jobs (17%). In terms of ethnicity, the sample consisted of 86% Whites, 5% African Americans, 3% Hispanics, and 3% Asians.

**Belgian sample.** As a similar survey panel system was not available in Belgium, research assistants tried to follow the participant recruitment procedure of the U.S. sample as closely as possible. Thus, 301 participants were also recruited from a variety of organizations in the Dutch-speaking part of Belgium. To this end, research assistants contacted the organizations via mail or by telephone. We tried as best as we could to include organizations from different regions. If the organization agreed to participate, we sent questionnaires accompanied by a reference letter to the contact person in the organization, who in turn was responsible for distributing these materials further. This sample consisted of Dutch-speaking Belgians (39% females and 61% males). This was a response rate of about 60%. The mean age of the respondents was 40.5 years. Their mean work experience was 15.2 years. The two most dominant job categories were technical jobs (31%) and clerical jobs (17%).

**Cyprus sample.** These participants were also recruited from a variety of organizations in the Greek-speaking part of Cyprus. The same procedure as in the Belgian sample was used. The response rate was also about 60%. This third sample consisted of 132 Greek-speaking Cypriots (59% females and 41% males). The mean age of the respondents was 35.1 years. Their mean work experience was 7.6 years. The two most dominant job categories were sales-related jobs (31%) and clerical jobs (19%).

**Measure and Translation Procedure**

All respondents completed the 18-item PSQ in their language, indicating their degree of satisfaction with various facets of pay on a 5-point scale, ranging from 1 (very dissatisfied) to 5 (very satisfied). We asked professional interpreters to translate the complete PSQ into Dutch and Greek. Next, professional interpreters back translated the questionnaires to English. Finally, six native English speakers (five men and one woman) compared the back translation to the original English version. They were asked to indicate whether the meaning of each sentence (item) had
remained similar on a 5-point scale ranging from 1 (the meaning of the sentence has not remained similar) to 5 (the meaning of the sentence has remained similar). On average, the PSQ items obtained a rating of 3.97 ($SD = 1.07$). There were two items that obtained a rating below 3, indicating that the translation was considered to be somewhat different from the back translation. Two of the authors scrutinized these items and made changes accordingly.

**Results**

**Test of Fit of Measurement Models Underlying the PSQ (Within Each Group)**

Preliminary study of the data showed no serious problems with nonnormality. All variables had skewness values $< 1.66$ and kurtosis values $< 1.42$. These values are well below the guidelines of severe normality (skewness values $> 2$ and kurtosis values $> 7$) suggested by Curran, West, and Finch (1996). Hence, we could use maximum likelihood (ML) as the estimation method in the confirmatory factor analyses.

We began by testing several measurement models that represented different conceptualizations of the PSQ. First, we tested a one-factor model. This model represented the early view of pay satisfaction as a unitary construct (e.g., Motowidlo, 1982). Second, we tested a five-factor model. This model specified pay satisfaction as a multidimensional construct, consisting of five different facets—pay level, pay raise, benefits, pay administration, and pay structure—as separate dimensions underlying pay satisfaction. This five-factor model was originally conceived by H. G. Heneman and Schwab (1985). Finally, we tested a four-factor model. This is the same model as the previous one, with the pay administration and pay structure scales collapsed. This model has received the most empirical support in prior research (e.g., DeConinck et al., 1996; R. L. Heneman et al., 1988; Judge, 1993; Judge & Welbourne, 1994).

To test the fit of these measurement models through CFA within each country, we used EQS 6 (Bentler & Wu, 2002) to derive ML estimates for the input covariance matrix. Besides the $\chi^2$ and the $\chi^2/df$, we used several fit indices to assess how these models represented the data. Our choice of absolute and incremental fit indices was based on Hu and Bentler (1999). In particular, on the basis of their recommendations, we used the Tucker-Lewis index (TLI), the comparative fit index (CFI), the standardized version of the root mean squared residual (SRMR), and the root mean square error of approximation (RMSEA). In terms of cutoff values, Hu and Bentler (1999) showed that a cutoff value close to .95 should be used for the ML-based TLI and CFI, a cutoff value close to .08 for SRMR, and a cutoff value close to .06 for the RMSEA.

Results of the confirmatory factor analyses by country are presented in Table 1. Both the four-factor and the five-factor model yielded a reasonably good fit to the data (especially in Belgium and Cyprus) because the fit indices of both of these
models were close to the aforementioned cutoff values. However, further inspection of the parameter estimates of the five-factor model showed that Factor II (Pay Administration) and Factor III (Pay Structure) could not be empirically distinguished from each other. For instance, in the Belgian sample, the correlation between these two factors was .96, in the Cyprus sample it was .82, and in the U.S. sample it was even 1.00 (constrained to a higher bound by EQS). This indicates that it is better to collapse those two factors, resulting in a more parsimonious four-factor model as the best representation of the data.

Table 2 presents descriptive statistics and Cronbach $\alpha$s for the scores on the four scales of the PSQ. The 95% confidence intervals of the Cronbach $\alpha$s are also shown. These confidence intervals were computed on the basis of the procedure outlined in Fan and Thompson (2001). With one exception, the $\alpha$ coefficients obtained in this study exceed the .80 criterion deemed appropriate for this type of research (see Henson, 2001).

Tests of Invariance of the Measurement Model (Stacked Multiple Groups)

Once a baseline model was established within each country, we examined the invariance of this model across the three countries. We conducted the following
sequence of increasingly more restrictive tests of invariance across groups via EQS 6 (see Byrne, 1998; Hancock, 1997): (a) factor form (i.e., the same number of factors and the factors have the same variables that load on them), (b) factor pattern coefficients, and (c) factor variances and covariances. Note that we did not test the equality of the error variances because this is generally considered to be an overly restrictive test (Byrne, 1998; Hittner, 1995; Yin & Fan, 2003).

To determine whether constraining parameters to be invariant across groups yielded a meaningful decrease in fit, the $\Delta \chi^2$ has traditionally been used as the index of difference in fit. However, the use of $\Delta \chi^2$ has been criticized because of its sensitivity to sample size (Cheung & Rensvold, 2002; Kelloway, 1995). Recently, Cheung and Rensvold (2002) provided evidence that $\Delta$CFI was not prone to these problems. On the basis of extensive simulations, they also determined that a $\Delta$CFI value higher than .01 was indicative of a meaningful drop in fit. If the $\Delta$CFI indicated that the constrained model did not lead to a meaningful decrease in fit as compared to the unconstrained model, the constrained parameters were considered to be invariant across groups.

Table 2
Descriptive Statistics and Cronbach’s $\alpha$s of Pay Satisfaction Questionnaire Scales Within Each Country

<table>
<thead>
<tr>
<th></th>
<th>United States ($N = 321$)</th>
<th>Belgium ($N = 301$)</th>
<th>Cyprus ($N = 132$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>2.72</td>
<td>3.02</td>
<td>3.08</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.02</td>
<td>.83</td>
<td>1.07</td>
</tr>
<tr>
<td>Skewness</td>
<td>.09</td>
<td>-.42</td>
<td>-.25</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-.88</td>
<td>-.35</td>
<td>-.79</td>
</tr>
<tr>
<td>Cronbach’s $\alpha$</td>
<td>.94</td>
<td>.83</td>
<td>.87</td>
</tr>
</tbody>
</table>

Note: CI = confidence interval.
Table 3 presents the results of the sequence of increasingly more restrictive tests of measurement invariance. As a first step, we examined whether the factor form of the measurement underlying the PSQ was equivalent across the three countries. Form invariance (also known as configural invariance) implies that observed scores relate to the same number of factors and that the factors have the same variables that load on them. If the factor structures underlying respondents’ ratings are the same across countries, this indicates that respondents use a similar frame of reference (conceptual domain) when rating the PSQ items (Riordan & Vandenberg, 1994). Table 3 shows that when we constrained the factor form, a good fit was obtained for this model, $\chi^2(387, N = 752) = 950.86$, CFI = .95, and RMSEA = .04.

So we continued with our tests of measurement invariance and constrained the factor pattern coefficients to be invariant across samples. Invariance of factor pattern coefficients (also known as metric invariance) implies that respondents calibrate the intervals used on the measurement scale in similar ways (Riordan & Vandenberg, 1994). As shown in Table 3, this additional set of constraints did not produce a meaningful drop in fit. In fact, $\Delta$CFI was not higher than .01, thereby not exceeding the critical value of Cheung and Rensvold (2002).

Finally, we tested for the invariance of factor variances and covariances across the samples. This means that we examined whether each of the four PSQ factors showed equal variance across the samples and whether the interrelations among these factors were the same. This test of invariance was supported because the $\Delta$CFI was not higher than .01.

**Discussion**

This study investigated the measurement invariance across three countries that are considered to be culturally different from each other (United States, Belgium, and Cyprus). Results showed that the measurement structure of the PSQ was invariant across these different countries. Specifically, there was no meaningful departure
from measurement invariance in terms of factor form, factor pattern coefficients, factor variances, and factor covariances. These results bode well for the equivalence of PSQ ratings across different countries.

Future research should further test the equivalence of the PSQ across different contexts. Although the present study supports the stability of the PSQ factor structure across three countries, previous studies have indicated that contextual variables might affect the factor structure obtained (e.g., Scarpello et al., 1988). Therefore, future studies should examine the equivalence of the PSQ across different countries, samples (e.g., occupational groups), and contexts (e.g., market types, industries) to understand better the effect of contextual variables on the factor structure of the PSQ. As another avenue of future research, more support for the multidimensionality of pay satisfaction should be obtained from studies focusing on differential prediction issues.

Finally, this study focused on four well-established dimensions of pay satisfaction. However, in recent years, compensation practices have dramatically changed. Several new facets of individuals’ pay packages may determine employees’ pay satisfaction. For instance, recent studies have suggested that lump-sum bonuses (Sturman & Short, 2000) and group incentive plans (Fong & Shaffer, 2003) constitute new and distinct dimensions of pay satisfaction. Future studies should test the validity of these new pay satisfaction dimensions.

References


