

Jang, S., Kim, E.S., Cao, C., Allen, T.D., Cooper, C.L., Lapierre, L.M., O'Driscoll, M.P., Sanchez, J.I., Spector, P.E., Poelmans, S., Abarca, N., Alexandrova, M., Antoniou, A-S., Beham, B., Brough, P., Crikci, I., Ferreira, P., Fraile, G., Geurts, S., Kinnunen, U., Lu, C.-Q., Lu, L., Moreno-Vela ´zquez, Y.F., Pagon, M., Pitariu, H., Salamatov, H., Siu, O.-L., Shima, S., Schulmeyer M.K., Tillemann, K., Widerszal-Bazyl, M., & Woo, J.-M. (in press). Measurement invariance of the Satisfaction with Life Scale across 26 countries. *Journal of Cross-Cultural Psychology*. Accepted 31/1/17.

Measurement Invariance of the Satisfaction with Life Scale Across 26 Countries

Seulki Jang¹, Eun Sook Kim¹, Chunhua Cao¹, Tammy D. Allen¹,
 Cary L. Cooper², Laurent M. Lapierre³, Michael P. O'Driscoll⁴, Juan I. Sanchez⁵, Paul E.
 Spector¹, Steven A. Y. Poelmans⁶, Nureya Abarca⁷, Matilda Alexandrova⁸, Alexandros-
 Stamatios Antoniou⁹, Barbara Beham¹⁰, Paula Brough¹¹, Ilker Carikci¹², Pablo Ferreira¹³,
 Guillermo Fraile¹⁴, Sabine Geurts¹⁵, Ulla Kinnunen¹⁶, Chang-qin Lu¹⁷, Luo Lu¹⁸, Ivonne F
 Moreno-Vela ´zquez¹⁹, Milan Pagon²⁰, Horea Pitariu²¹, Volodymyr Salamatov²², Oi-ling Siu²³,
 Satoru Shima²⁴, Marion K Schulmeyer²⁵, Kati Tillemann²⁶, Maria Widerszal-Bazyl²⁷, & Jong-
 Min Woo²⁸

¹ University of South Florida, Tampa, USA; ² Lancaster University, Lancaster, UK; ³ University of Ottawa, Ottawa, Canada; ⁴ University of Waikato, Waikato, New Zealand; ⁵ Florida International University, Miami, USA; ⁶ EADA Business School, Barcelona, Spain; ⁷ Pontificia Universidad Cato ´lica de Chile, Santiago, Chile; ⁸ University of National and World Economy, Sofia, Bulgaria; ⁹ Athens University, Athens, Greece; ¹⁰ Humboldt-Universita ´t zu Berlin, Berlin, Germany; ¹¹ Griffith University, Brisbane, Australia; ¹² Suleyman Demirel University, Isparta, Turkey; ¹³ Universidad de Piura, Lima, Peru; ¹⁴ Austral University, Buenos Aires, Argentina; ¹⁵ Radboud University Nijmegen, Heijendaal, Netherlands; ¹⁶ University of Tampere, Tampere, Finland; ¹⁷ Peiking University, Beijing, China; ¹⁸ National Taiwan University, Taipei, Taiwan; ¹⁹ University of Puerto Rico, Rio Piedras Campus, San Juan, USA; ²⁰ Al Ghurair University, Dubai, United Arab Emirates; ²¹ Babes-Bolyai University, Cluj-Napoca, Romania; ²² National Academy of Public Administration, Kiev, Ukraine; ²³ Lingnan University, Hong Kong, China; ²⁴ Kyoto Bunkyo University, Uji, Kyoto, Japan; ²⁵ Private University of Santa Cruz of the Sierra, Santa Cruz, Bolivia; ²⁶ Estonian Business School, Tallinn, Estonia; ²⁷ Central Institute for Labor Protection, Warsaw, Poland; ²⁸ Inje University, Seoul Paik Hospital, Seoul, South Korea

Abstract

The Satisfaction with Life Scale (SWLS) is a commonly used life satisfaction scale. Cross-cultural researchers use SWLS to compare mean scores of life satisfaction across countries. Despite the wide use of SWLS in cross-cultural studies, measurement invariance of SWLS rarely has been investigated and previous studies showed inconsistent findings. Therefore, we examined the measurement invariance of SWLS with samples collected from 26 countries. In order to test measurement invariance, we utilized three measurement invariance techniques: (1) Multi-Group Confirmatory Factor Analysis (MG-CFA), (2) Multi-Level Confirmatory Factor Analysis (ML-CFA), and (3) alignment optimization methods. The three methods demonstrated that configural and metric invariances of life satisfaction held across 26 countries, while scalar invariance did not. With partial invariance testing, we identified that the intercepts of items 2, 4, and 5 were non-invariant. Based on two invariant intercepts, factor means of countries were compared. Chile showed the highest factor mean; Spain and Bulgaria showed the lowest. The findings enhance our understanding of life satisfaction across countries, and they provide researchers and practitioners with practical guidance on how to conduct measurement invariance testing across countries.

Measurement Invariance of the Satisfaction with Life Scale Across 26 Countries

Life satisfaction is referred to as “a global assessment of a person’s quality of life according to his chosen criteria” (Shin & Johnson, 1978, p.478), and it has been identified as a significant subjective indicator of well-being (e.g., Andrews & Whitney, 1976; Diener, Emmons, Larsen, & Griffin, 1985). Different from emotional aspects of well-being (positive affect and negative affect), life satisfaction involves a cognitive judgment of one’s life (Andrews & Withey, 1976). Life satisfaction is a cognitive comparison of one’s current state with one’s standard of what is appropriate or desirable (Diener, 1984). If one’s current state matches one’s standard, then the person will be likely to experience a high level of life satisfaction.

Life satisfaction has been frequently measured with the Satisfaction with Life Scale (SWLS; Diener et al., 1985). SWLS has been found to be a reliable and valid measure. For example, SWLS has high reliability (higher than $\alpha = .80$; Pavot, Diener, Colvin, & Sandvik, 1991), and studies support convergent validity with other life satisfaction scales and subjective well-being measures (Pavot et al., 1991), concurrent validity with health (Lyubomirsky, King, & Diener, 2005) and predictive validity on suicide attempts (Koivumaa-Honkanen, Honkanen Viinamaeki, Heikkilae, Kaprio, & Koskenvuo, 2001).

Based on the SWLS measure, cross-cultural researchers have investigated life satisfaction in individual countries and compared life satisfaction scores across countries (e.g., Oishi, Diener, Lucas, & Suh, 1999). However, people from different countries have varying cultural values and practices (e.g., GLOBE study; House, Hanges, Javidan, Dorfman, & Gupta, 2004), which might affect respondents’ understanding and reporting of life satisfaction. Consequently, they may perceive life satisfaction items differently, and interpret response scales differently. Such factors can contribute to nonequivalence of measures, or measurement non-invariance, which means the

construct is interpreted in a different way across countries. If that is the case, the same mean scores of samples from different countries may not indicate the same actual levels of life satisfaction. Correlation coefficients and regression coefficients can also be biased and misleading (e.g., Reise, Widaman, & Pugh, 1993). Therefore, it is critical to examine measurement invariance before conducting group comparisons in terms of relations (e.g., correlations) and means.

Despite the criticality of measurement invariance testing, the majority of life satisfaction studies using SWLS have omitted measurement invariance testing before conducting mean comparisons (e.g., Oishi et al., 1999). In addition, the handful of studies that examined measurement invariance of SWLS across countries compared only two or three countries and showed inconsistent findings (e.g., Eid, Langeheine & Diener, 2003; Hofer, Chasiotis, & Campos, 2006; Oishi, 2006). For instance, Eid and colleagues (2003) performed a latent class analysis to test measurement invariance of SWLS between the United States (U.S.) and China, finding that measurement invariance did not hold. Specifically, Chinese participants were more modest about reporting greater life satisfaction. Because the Chinese tend to endorse less extreme responses than do those from the U.S. (e.g., Roster, Albaum, & Rogers, 2006), different response styles could play a role in measurement non-invariant results.

Similarly, Oishi (2006) examined measurement invariance of SWLS between students from the U.S. and Chinese students, using Multi-Group Structural Equation Modeling, Multiple Indicators Multiple Causes, and Item Response Theory techniques. All three techniques consistently found that item 4 (“So far I have gotten the important things I want in life”) and item 5 (“If I could live my life over, I would change almost nothing”) were non-invariant, and Chinese students endorsed these two items less than did U.S. students. Oishi argued that items 4

and 5 measure satisfaction with previous accomplishments while the other three items measure satisfaction with present conditions. Because East Asians tend to underrate their previous accomplishments or performance as a sign of modesty (e.g., Heine, Lehman, Markus, & Kitayama, 1999, for review), they are less likely to endorse items 4 and 5. Also, China is a self-critical society where continuous self-improvement is valued and standards are getting higher over time for continuous self-improvement (Markus & Kitayama, 1991). Therefore, Chinese respondents might not be satisfied with their past accomplishments based on their newer and higher standard, leading to less endorsement on items 4 and 5.

Hofer et al. (2006) stated that they tested for and found support for measurement invariance of SWLS across three countries (Costa Rica, Cameroon, and Germany). However, they failed to use a proper measurement invariance testing technique such as MG-CFA; instead, they combined all samples from the three countries and performed regular CFA to examine measurement invariance. Due to this inappropriate testing method, their conclusions seem unreliable.

Based on the limited research to date, it is difficult to draw a general conclusion on measurement invariance of SWLS across countries. Therefore, we examined the measurement invariance of SWLS across 26 countries and used three different measurement invariance techniques: (1) Multi-Group Confirmatory Factor Analysis (MG-CFA; e.g., Millsap, 2011), (2) Multi-Level Confirmatory Factor Analysis (ML-CFA; e.g., Jak, Oort, & Dolan, 2013), and (3) alignment optimization (Asparouhov & Muthén, 2014). These methods were selected for three reasons. First, MG-CFA is the most frequently and conventionally used method. Second, ML-CFA takes into account the multi-level nature of the data structure in cross-cultural research. Lastly, alignment optimization has proposed to be appropriate when a large number of groups

are compared. It should be noted that there are other approaches to testing measurement invariance across a large number of groups (e.g., Bayesian approximate measurement invariance testing, multilevel factor mixture modeling); however, we limit our study to these three methods.

The first purpose of this paper is to test whether the measurement invariance of SWLS holds across 26 countries. The second purpose is to introduce and compare three different measurement invariance techniques. Asparouhov and Muthén (2014) argued that MG-CFA has significant limitations in measurement invariance testing when a large number of groups are compared. Hence, we included two alternative methods to test measurement invariance in addition to MG-CFA. We compared their testing procedures as well as their results.

***Research Question 1:** Does the Satisfaction with Life Scale (SWLS) show measurement invariance across 26 countries?*

***Research Question 2:** Do different measurement invariance testing methods show consistent results in measurement invariance across 26 countries?*

In the next section, the three measurement invariance techniques are described in detail.

Measurement Invariance

Measurement invariance or equivalence refers to “lack of bias” (Meredith & Millsap, 1992, p. 209) and tests whether “measurements yield measures of the same attributes (Horn & McArdle, 1992, p. 117). It has been recognized as a crucial step for group comparison studies as it demonstrates whether different group members interpret the survey items in the same way with similar response anchors (e.g., Vandenberg & Lance, 2000). Moreover, it allows researchers to compare different groups in a meaningful way with respect to their means and correlations between variables (Vandenberg & Lance, 2000; Cheung & Rensvold, 2002).

Measurement invariance is typically tested at four levels incrementally (Horn & McArdle,

1992). To be specific, configural invariance examines whether items load onto the same latent factor across groups; however, factor loadings, intercepts, and residual variances are freely estimated (Horn, McArdle, & Mason, 1983). If configural invariance holds, it indicates that the latent structure is similar across groups. Once configural invariance holds, metric invariance is tested. Metric invariance means that the factor loading of each item on the latent factor is the same across groups. Satisfying metric invariance demonstrates that the unit and the interval of the latent factor are equal across groups (Chen, 2007). Thus, it allows the comparison of factor variances and structural relations (e.g., correlations between variables) across groups (Asparouhov & Muthén, 2014). Once metric invariance holds, scalar invariance is tested to examine whether the intercept of each item is the same across groups in addition to the equality of factor loadings. Importantly, meeting scalar invariance allows researchers to compare latent factor means, latent factor variances, and relevant covariance between groups (Meredith, 1993). Lastly, strict invariance can be tested to investigate whether the residual variance of each item is the same across groups in addition to the equality of factor loadings and intercepts. Meeting strict invariance provides confidence that the group mean differences on the scale scores are driven from real group differences and not from other factors. However, scalar invariance is considered sufficient to meaningfully compare factor or observed means (Meredith, 1993). In this study, MG-CFA, ML-CFA, and alignment optimization methods were used to investigate measurement invariance of the SWLS across 26 countries.

(1) Multi-Group Confirmatory Factor Analysis (MG-CFA)

The most frequently used measurement invariance testing technique is a multi-group confirmatory factor analysis (MG-CFA; e.g., Millsap, 2011). The ultimate purpose of MG-CFA is to compare latent factor means, latent factor variances, and relevant covariance between

groups after controlling for measurement errors. MG-CFA usually treats groups as a fixed classification. In other words, particular groups in a study (e.g., gender) are considered as all possible groups in the population.

Although MG-CFA is the most well-established method, MG-CFA is cumbersome and impractical when many groups are compared (e.g., Asparouhov & Muthén, 2014). Also, model fit indices (e.g., Chi-square, CFI, RMSEA) may not perform reasonably in multiple-group comparisons, and considerable modifications may be needed to improve model fit at the scalar level, which possibly leads to a higher chance of incorrect model specification (Asparouhov & Muthén, 2014). To address these limitations, in addition to MG-CFA we also used two alternative methods, ML-CFA and alignment optimization.

(2) Multi-Level Confirmatory Factor Analysis (ML-CFA)

The first alternative we adopted is a multi-level confirmatory factor analysis (ML-CFA; e.g., Jak et al., 2013). ML-CFA treats groups as a random sample from the population (e.g., 20 countries are randomly selected from all countries in a region of interest). ML-CFA is a combination of multilevel models (accounting for the hierarchical structure of individuals nested in group units) and structural equation modeling (taking into account measurement errors). ML-CFA decomposes the total variance into two components (i.e., within country variance and between country variance) and thus allows researchers to construct a measurement (or CFA) model at both individual-level and country-level (within-level and between-level, interchangeably) using within-country and between-country variance covariance matrices. Similar to measurement invariance in MG-CFA, ML-CFA incrementally tests configural, metric, and scalar measurement invariances across groups or clusters.

Configural invariance is tested by specifying the same factor model for within-level and

between-level comparisons. Good multilevel model fit indicates configural invariance. Once configural invariance is satisfied, metric invariance is tested by constraining within-level CFA factor loadings and between-level CFA factor loadings to be equal. The rationale behind this constraint is that if factor loadings are invariant across all groups, both within-level CFA factor loadings and between-level CFA factor loadings should be identical. Once metric invariance holds, scalar invariance is tested by constraining the between-level residual variances to 0. The reason behind this constraint is that when intercepts of all groups are identical, the variability of intercepts across groups is 0; that is, the between-level residual variances should be 0. Jak et al. (2013) provide the mathematical proof of metric and scalar invariances across clusters.

(3) Alignment Optimization Methods

The second alternative to MG-CFA is alignment optimization (Asparouhov & Muthén, 2014). Alignment optimization searches for the most optimal measurement invariance. That is, it finds the most prominent non-invariance in a small number of items allowing most of the items to have a minimal amount of difference in intercept and loading parameters. Alignment optimization estimates factor mean and variance parameters within each group to minimize the total amount of non-invariance, instead of automatically assuming measurement invariance. Requirements for the most optimal solution are the minimized number of non-invariance parameters and the minimized amount of non-invariance. Unlike MG-CFA and ML-CFA that test the four levels of measurement invariance stepwise, alignment optimization examines the invariance of factor loadings and intercepts simultaneously. In alignment optimization, invariance is tested by fixing the *factor mean* of each group (α_g) to 0 and the *factor variance* of each group (ψ_g) to 1. All loadings and intercepts are freely estimated (i.e., configural invariance). Then, factor means and variances of each group are computed with the approximate invariance

assumption. Compared to the exact invariance assumption (i.e., factor loadings and intercepts are identical across all groups), the approximate invariance assumption is less stringent, especially when a large number of groups are compared. Alignment optimization attempts to minimize non-invariance instead of constraining factor loadings and intercepts to be equal across groups to estimate the factor means and variances of the groups. Specific procedures of measurement invariance testing and technical details such as the computation of the total loss function and the component loss function are thoroughly described in Asparouhov and Muthén (2014).

Two alignment optimization methods can be used: FIXED optimization and FREE optimization. In FIXED optimization, the factor mean and the factor variance of the first group are fixed to 0 and 1, respectively; in FREE optimization, there is no constraint on the first group's factor mean and variance and they are freely estimated. Two types of estimators can be used: maximum likelihood (ML) and Bayesian alignment estimation. ML relies on asymptotic theory, while the Bayes method relies on prior specifications (Asparouhov & Muthén, 2014). In order to use the Bayes method, researchers need to specify the distributions of parameters in the model (i.e., priors) based on empirical research results. In this study, because we had little information about the distributions of factor loadings and intercepts across 26 countries, ML was chosen as the estimation method.

Contributions

The current study contributes to the existing literature in five ways. First, this study provides evidence to help resolve inconsistent findings concerning the measurement equivalence of the SWLS in previous research. While some studies found a lack of scalar invariance (e.g., Hofer et al., 2006), others reported scalar invariance (e.g., Eid et al., 2003). Therefore, we include a large number of countries, and investigate the measurement invariance of life

satisfaction. Accordingly, this study provides results that are likely more generalizable than previous studies. Second, the use of the three measurement invariance testing methods helps provide robust conclusions. Third, this study offers information about non-invariant items and non-invariant countries beyond general model fit information. Fourth, we conduct partial scalar invariance testing and compare the factor means of 26 countries. Lastly, we offer practical guidance for researchers and practitioners on how to conduct measurement invariance testing across a large number of countries.

Method

Procedures and Participants

The present study uses data from the second phase of the Collaborative International Study of Managerial Stress (CISMS 2; Spector et al., 2007). CISMS 2 data was collected from approximately 2003 to 2005. Participants were 7004 managers from local companies in 26 countries. Average within country sample size was 268 and ranged from 137 (United Kingdom) to 500 (Australia). Of the 7004 managers, 61% were male and average age was 39.80 ($SD = 10.44$). Specific demographic information for each country appears in the online supplement.

The survey was designed by a central data collection team comprised of researchers in psychology and in organizational behavior. For countries in which English was not a main language, the survey was translated into the dominant language by a research in that country. U.S. doctoral students independently performed back translations (van de Vijver & Leung, 1997). If there was disagreement on a translation, the back translators modified the translation. Few translation errors were found. In total, SWLS was translated into 15 languages.

Measures

Life Satisfaction. All five items of SWLS from Diener et al. (1985) were included with

response options that ranged from 1 (strongly disagree) to 7 (strongly agree).² Item 1 is “In most ways my life is close to my ideal”; Item 2 is “The conditions of my life are excellent”; Item 3 is “I am satisfied with my life”; Item 4 is “So far I have gotten the important things I want in life”; and Item 5 is “If I could live my life over, I would change almost nothing.” Our data showed a high Cronbach’s α ($\alpha = .90$) across all respondents. Each country’s Cronbach’s α was above .80, except Bulgaria ($\alpha = .60$). Specific α values for each country are in the online supplement.

Results

Preliminary Data Analysis

Before measurement invariance is tested across countries, it is important to ensure sample invariance on demographic variables (Douglas & Craig, 1983). Evidence of satisfying sample invariance provides confidence that obtained results are attributed to country differences, not to demographic dissimilarities across samples. Specifically, age invariance was investigated using a series of one-way analyses of variance (ANOVA). Previous studies revealed that SWLS was non-invariant across different age groups though SWLS was invariant across gender (e.g., Hultell & Gustavsson, 2008). Hultell and Gustavsson (2008) found that older respondents endorsed higher scores on item 4 (“So far I have gotten the important things I want in life.”) and item 5 (“If I could live my life over, I would change almost nothing.”) than younger respondents. They argued that older respondents tend to be more established, experienced more life events, and made more critical life decisions than younger respondents; therefore, older respondents are likely to report higher scores on the two items. Based on previous findings, we specifically tested invariance of age across various country samples using one-way ANOVA.

² United States, United Kingdom, and Canada data were collected using a five-Likert scale that ranged 1 (strongly disagree) to 5 (strongly agree). Dawes (2008) demonstrated that five-point scores are comparable to seven-point scores and can be converted to seven-point scores without significant changes on psychometrical properties. Based on Dawes’s (2008) suggestion, the five-point data were converted to the seven-point data in this current study. Psychometrical properties between the five-point data and the seven-point data were almost identical.

One-way ANOVA revealed that age composition was significantly different across the 26 countries [$F(25, 6943) = 65.63, p < .05$]. This indicates that participants from at least one country were significantly older or younger than those from the other countries. The different age compositions across countries might affect measurement invariance of SWLS.

Measurement Invariance

MG-CFA, ML-CFA, and alignment optimization methods were conducted to investigate measurement invariance of SWLS across 26 countries. We specified a one-factor model using Mplus 7.2. All variables were treated as continuous because SWLS is a 7-point Likert scale (agree-disagree). A substantial body of literature suggested that treating Likert scale variables as continuous is acceptable when the number of response categories is five or more (e.g., Marsh et al., 2013). Also, categorical estimation procedures (e.g., weighted least squares) do not necessarily produce better solutions due to computational limitations under current hardware developments particularly for large and complex models (Marsh et al., 2013, p. 116). Responses in the current data were approximately normally distributed in terms of kurtosis and skewness.

(1) Multi-group Confirmatory Factor Analysis (MG-CFA)

In this study, configural, metric, and scalar invariances were tested. However, strict invariance was not tested because this approach is too stringent and unrealistic (Byrne, 1994). In MG-CFA, a maximum likelihood estimator was used by default and the country variable was regarded as a grouping variable. For the identification of the configural invariance model, the factor loading of the first item was fixed to one and the intercept of this item was constrained to be equal across countries. Residual covariances of all indicators were fixed to zero. Syntax is provided in the online supplement.

Results were interpreted based on the three fit indices: Chi-square, comparative fit index

(CFI), and root-mean-square error of approximation (RMSEA). However, chi-square values are sensitive to sample sizes and to large number of groups (e.g., Rutkowski & Svetina, 2014).

Therefore, CFI and RMSEA are regarded as more robust indicators. Conventionally, a CFI higher than .95 and a RMSEA below .05 indicate good model fit (Hu & Bentler, 1999). However, when many groups are included, RMSEA tends to become greater than .05 regardless of actual model fit (e.g., Rutkowski & Svetina, 2014). Therefore, Rutkowski and Svetina (2014) suggested a more liberal RMSEA cutoff (.10) for at least 10 groups. Currently, there is no specific cutoff suggested for more than 20 groups. However, an even more liberal RMSEA cutoff may be adopted given that Rutkowski and Svetina's (2014) simulation showed that number of groups made an impact on the RMSEA cutoff (the more, the higher) and the estimated RMSEA values were over .15 when non-invariance was present. Thus, the current results from 26 country groups were interpreted based on a RMSEA cutoff of .15. As for model fit comparison indices, $\Delta\text{CFI} \leq .020$ and $\Delta\text{RMSEA} \leq .030$ were used in evaluating metric invariance from configural invariance; $\Delta\text{CFI} \leq .010$ and $\Delta\text{RMSEA} \leq .015$ were used in evaluating scalar invariance from metric invariance (Rutkowski & Svetina, 2014).

First, a configural invariance model with one factor was evaluated. Results showed acceptable fit to the data, [$\chi^2(130) = 827.860, p < .01, \text{CFI} = .964, \text{RMSEA} = .141$]. The chi-square test was statistically significant ($p < .01$), but because it is possible that the null hypothesis was falsely rejected due to the large sample size in this study, we further investigated other fit criteria. The CFI supported good model fit. We considered the RMSEA acceptable with a more lenient cutoff (.15) based on Rutkowski and Svetina's (2014) suggestion. Considering the reasonable overall model fit based on CFI and RMSEA values, we concluded there was support for configural invariance. Second, metric invariance was tested. Results showed that the model

fit the data adequately, [χ^2 (230) = 1215.898, $p < .01$, CFI = .949, RMSEA = .126]. The model comparison test (configural vs. metric) suggested metric invariance, [$\Delta\chi^2$ (100) = 388.038, $p < .01$, Δ CFI = .015, Δ RMSEA = -.015]. Although the chi-square difference test was significant ($\Delta\chi^2$ (100) = 388.038, $p < .01$), considering the adequate overall model fit and the support of model comparison fit indices (Δ CFI and Δ RMSEA), we concluded that metric invariance holds. Lastly, scalar invariance was tested. The model failed to fit the data, [χ^2 (330) = 2687.878, $p < .01$, CFI = .877, RMSEA = .163]. A model comparison test was conducted and revealed that scalar invariance did not hold, [$\Delta\chi^2$ (100) = 1471.98, $p < .01$, Δ CFI = .072, Δ RMSEA = .037]. Taking all these results into consideration, scalar invariance was not satisfied across the 26 countries.

Due to the lack of scalar invariance, partial scalar invariance was conducted on the basis of metric invariance (Byrne, Shavelson, & Muthen, 1989). Partial invariance is a compromised way to handle the lack of invariance because it relaxes non-invariant items while invariant items are still constrained. One advantage of partial invariance testing is to allow researchers to compare factor means without satisfying full measurement invariance as long as at least two items are invariant. In the baseline model, item 3 was selected as a reference item (its intercept was constrained to be equal across all countries) because item 3 was found to have the smallest residual variance across countries in metric invariance.³ Next, the intercept of item 1 was constrained to be equal across countries on the basis of the metric invariance baseline model, and model fit was compared against the baseline model. This same process was repeated for items 2, 4, and 5, separately. Results showed that partial scalar invariance constraining item 1 across countries held [CFI = .949, RMSEA = .126, Δ CFI = .010, Δ RMSEA = .004]. However, partial scalar invariance constraining item 2 across countries did not hold [CFI = .935, RMSEA = .135,

³To confirm this decision, we also tested partial scalar invariance constraining the intercept of item 3 to be equal across countries while item 1 was a reference item. Results showed that the partial scalar invariance constraining item 3 across countries held [CFI = .949, RMSEA = .126, Δ CFI = .010, Δ RMSEA = .004].

$\Delta\text{CFI} = .014$, $\Delta\text{RMSEA} = .009$]; partial scalar invariance constraining item 4 across countries did not hold [$\text{CFI} = .932$, $\text{RMSEA} = .137$, $\Delta\text{CFI} = .017$, $\Delta\text{RMSEA} = .011$]; and partial scalar invariance constraining item 5 across countries did not hold [$\text{CFI} = .928$, $\text{RMSEA} = .142$, $\Delta\text{CFI} = .021$, $\Delta\text{RMSEA} = .016$]. Therefore, we concluded that partial scalar invariance testing constraining items 1 and 3 was satisfied. When the intercepts of items 1 and 3 were constrained while the intercepts of items 2, 4, and 5 were relaxed, results showed good model fit, [$\Delta\chi^2(25) = 202.937$, $p < .01$, $\text{CFI} = .939$, $\text{RMSEA} = .130$, $\Delta\text{CFI} = .014$, $\Delta\text{RMSEA} = .009$].

When partial scalar invariance shows at least two invariant items (i.e. one anchor item and one additional item), factor means can be meaningfully compared (Byrne et al., 1989; Steenkamp & Baumgartner, 1998). Because our partial scalar invariance tests found two invariant items (items 1 and 3), factor means of 26 countries were compared and are presented in Table 1. Chile had the highest factor mean score while Spain and Bulgaria had the lowest.

(2) Multi-level Confirmatory Factor Analysis (ML-CFA)

Before measurement invariance was examined, the intraclass correlations (ICCs) of the five items were checked to see how much variance exists at the country-level and whether a multi-level approach is appropriate. ICC values of the five items ranged from .154 to .227 across the 26 countries, which indicates that a substantial proportion of variance in each item was present at the country-level. Therefore, we proceeded with the ML-CFA approach.

For measurement invariance testing with ML-CFA, configural, metric, and scalar invariances were tested in order. A maximum likelihood with robust standard errors and chi-square (MLR) estimator was used by default. Country was regarded as a cluster variable (i.e., countries were the between-level unit of analysis). By default, the first item factor loading was fixed to 1 and all residual covariances were fixed to 0 both within-level and between-level. The

relevant syntax is provided in the online supplement. ML-CFA measurement invariance results were interpreted based on two fit indices: CFI and RMSEA. CFI and RMSEA cut-off values follow the basic CFA cut-off value guidelines.

First, configural invariance was tested. Results showed that the model fit the data acceptably well, [$\chi^2(10) = 321.897, p < .01, CFI = .959, RMSEA = .067$]. Second, metric invariance testing was performed by constraining factor loadings to be equal across levels (i.e., within-factor loadings are equal to between-factor loadings for all items). The overall model fit was adequate, [$\chi^2(14) = 368.882, p < .01, CFI = .954, RMSEA = .060$]. Lastly, scalar invariance was tested by constraining the between-level residual variance to zero. Results showed poor fit to the data, [$\chi^2(19) = 1533.981, p < .01, CFI = .802, RMSEA = .107$]. Thus, partial scalar invariance was tested. Using modification indexes from the scalar invariance model, we searched for the source of non-invariance (or misfit). Then, we relaxed the parameter of the largest modification index one at a time. We re-ran the model after relaxing the corresponding parameter and repeated this process until none of modification index values was substantial (Yoon & Kim, 2014). Modification indexes showed that the between-level residual variances of items 2, 4, and 5 were large; thus, we sequentially relaxed the between-level residual variances of items 2, 4, and 5 to freely estimate these residual variances. We also confirmed that the residual variances of these three items were statistically significantly different from zero in the scalar invariance model. This implies that the three items had non-invariant intercepts. Consistent with MG-CFA, item-level analyses revealed that the intercepts of items 1 and 3 were invariant. Therefore, the partial scalar invariance model in which the between-level residual variances of items 2, 4, and 5 were freed showed good model fit, [$\chi^2(16) = 509.800, p < .01, CFI = .935, RMSEA = .066$]. Between-level factor variance was statistically significantly different

from zero ($\Psi_B = .328, p < .01$), which suggested that the estimated life satisfaction factor means varied across countries. However, ML-CFA does not produce the factor mean of each country by default. Therefore, the rank order of factor means was not evaluated in ML-CFA.

(3) Alignment Optimization Methods

Alignment optimization is executed with the mixture analysis (TYPE = MIXTURE) and the KNOWNCLASS option (see online supplement for Mplus syntax). In other words, in alignment optimization, country is regarded as a latent class variable. Generally, a latent class refers to an unobserved or latent categorical variable (or group); however, because country membership is an observed grouping variable, we specified that country membership is known using the KNOWNCLASS option. The mixture analysis with the KNOWNCLASS option is analogous to a multiple group analysis, but has greater modeling flexibility (Asparouhov & Muthén, 2012). A one-factor model with five observed items was estimated with a maximum likelihood estimation method.

Asparouhov and Muthén (2014) suggested that a FREE approach works better than a FIXED approach when more than two groups are compared and non-invariance is expected. Because this study includes 26 diverse country groups, we initially used a FREE approach; however, this model produced a warning of untrustworthy standard errors, and a FIXED approach was suggested as a solution to deal with this problem (Asparouhov & Muthén, 2014). Therefore, a FIXED approach was adopted with the U.S. as the reference group with a factor mean 0. The U.S. was selected as the reference group because the SWLS was developed and has been predominantly used there (Diener, 1985; Fredrickson, Tugade, Waugh, & Larkin, 2003). Relevant syntax is provided in the online supplement.

Alignment optimization provides information about non-invariant items and non-

invariant countries per item. Alignment optimization was performed using the samples from all 26 countries. Table 2 demonstrates an invariance pattern with the two alignment fit indices: (1) a fit function contribution value and (2) an R-squared value (Asparouhov & Muthén, 2014). First, a fit function contribution value highlights a potential non-invariant item. When an absolute fit function contribution value is high, the item is likely to be non-invariant. Also, the sum of all fit function contribution values represents an optimized simplicity function value. However, what value should be regarded as high or low is an arbitrary decision. Relatively, the intercepts of items 4 and 5 showed higher absolute fit function contribution values than the intercepts of the other items, which indicates that items 4 and 5 were possibly more non-invariant than the other items. Second, R-square represents the proportion of variance in the estimated parameters (i.e., factor loadings and intercepts) explained by variation in the factor mean and factor variance across all groups. Therefore, R-square indicates the degree of invariance in factor loadings and intercepts across groups. A low R-square indicates non-invariance. In Table 3, the R-squares of items 4 and 5 were relatively lower, which suggests possible non-invariance in these items.

Table 3 shows non-invariant countries in each item factor loading and in each item intercept. When a group is parenthesized, this indicates that the parameter of that group is non-invariant. For example, (EE) in the item 1 intercept indicates that Estonia had a non-invariant factor intercept for item 1. In general, the total number of parentheses in intercepts was larger than the total number of parentheses in loadings, which suggests that the intercepts of the items were more non-invariant than the loadings of the items. Given that equal loadings represent metric invariance, and equal intercepts in addition to equal factor loadings represent scalar invariance, the results suggest that metric invariance might hold and scalar invariance might not hold across the 26 countries. Specifically, several countries (8 to 10) showed non-invariance in

the intercepts of items 2, 4, and 5. Taken together, scalar non-invariance was suspected, especially in the intercepts of items 2, 4, and 5.

Alignment optimization allows for a comparison of factor means across groups without meeting exact scalar invariance because alignment optimization assumes approximate invariance instead of exact invariance. However, how much non-invariance is considered as approximate invariance is still ambiguous and understudied in the literature. Based on a simulation study, Muthén and Asparouhov (2014) suggested that if less than 25% of parameter estimates are non-invariant, the results of alignment would be valid and interpretable. However, this rough rule of thumb should be taken with great caution until it is verified in multiple simulation studies and the rule should be used along with other indicators of non-invariance. In this study, less than 25% of all factor loading estimates were non-invariant; however, more than 25% of the intercept estimates for items 2, 4, and 5 were non-invariant. Because a large number of countries were identified as non-invariant in the intercepts of items 2, 4, and 5, a meaningful comparison of group means was still questionable. Nevertheless, because at least two items appeared fairly invariant (i.e., partial invariance), we report the rank order of the factor means from the alignment optimization and compare it to the rank from MG-CFA (see Table 1). In order to compare the rank from the partial scalar MG-CFA and the rank from the alignment optimization, we calculated the Spearman rank-order correlation. The correlation was .97, which implies that the two methods produced very similar ranks of factor means. Still, interpretation of the factor mean rank should be done with caution due to partial invariance.⁴

Post-hoc Analysis: Country Mean Age as a Source of Scalar Non-Invariance

⁴ We also calculated the mean composite scores of the 26 countries using the observed scores of all items, and ranked them (see Table 1). The Spearman rank-order correlation was .95 with the rank order from the partial scalar invariance in MG-CFA, and the correlation was .91 with the rank order from the alignment optimization.

MG-CFA, ML-CFA, and alignment optimization revealed that the intercepts of items 2, 4, and 5 are non-invariant across countries. Although it is unclear why the intercepts were non-invariant, one possibility is a dissimilar age mean across countries based on results from the preliminary data analysis. Therefore, we used Multilevel Multiple Indicators Multiple Causes (ML-MIMIC) modeling to investigate whether the dissimilar mean age across countries affected the intercepts of the non-invariant items. MIMIC modeling incorporates observed predictors or covariates in CFA to explain latent factors and/or observed indicators. When a predictor (e.g., age) explains a latent factor, it indicates that factor scores vary depending on the level of the predictor (e.g., the older, the higher/lower factor scores). When a predictor explains an observed indicator, this indicates non-invariance of the intercept of the indicator because the observed scores are not solely explained by the latent factor but also by age (intercept non-invariance in terms of age). For ML-MIMIC, the predictors can be introduced within-level, between-level, or both within- and between-levels to explain intercept non-invariance at each level or both levels.

We extended ML-CFA that was already completed to ML-MIMIC (Jak et al., 2011) to account for measurement non-invariance using within-level and between-level covariates. In other words, ML-MIMIC is used to explain why non-invariance is present. Aggregate country mean age was modeled as a between-level covariate while age itself was included as a within-level covariate. Note that our focal interest was non-invariant intercepts across countries at the between-level. The between-level latent variable and item 2 were regressed on the country mean age at the between-level to explain the intercept non-invariance of item 2 across countries. This analysis was repeated for items 4 and 5. The MLR estimator was used by default. Results showed that the different age mean of countries significantly affected the intercept of item 4 [$\gamma = .193, p < .01$] at the between-level (the country-level), while the country mean age did not

significantly affect the item 2 intercept [$\gamma = .011, p = .85$] nor the item 5 intercept [$\gamma = .110, p = .29$] at the between-level. The between-level residual variance of item 4 was still statistically significant ($0.065, t = 3.667, p < .01$), implying that the intercept difference across countries was not fully explained by the country mean age, and there could be other sources of non-invariance. Moreover, the country mean age did not account for the variability in the intercepts of items 2 and 5 across the 26 countries. Thus, future research is called for to further explore the sources of non-invariance in SWLS with more substantive country characteristic variables.

Discussion

Previous measurement invariance tests of SWLS across countries have produced inconsistent results. To help resolve the inconsistent findings, we included multiple countries and examined SWLS measurement invariance using three different methods (MG-CFA, ML-CFA, and alignment optimization). Results from all three methods consistently revealed that configural and metric invariances of SWLS held, but scalar invariance did not. Therefore, partial scalar invariance testing was conducted in MG-CFA and ML-CFA by relaxing the intercepts of non-invariant items. Both MG-CFA and ML-CFA tests indicated that the intercepts of items 2, 4, and 5 were non-invariant while the intercepts of items 1 and 3 were invariant. Based on partial scalar invariance with two invariant items, MG-CFA was used to further compare and rank factor means. Chile had the highest factor mean of life satisfaction, while Spain showed the lowest. Although the alignment optimization method does not allow partial invariance testing, it provides information on approximate measurement invariance (non-invariance) for groups and the rank of factor means across groups. Results of the approximate measurement invariance (non-invariance) showed that more non-invariant countries were identified on the intercepts of items 2, 4, and 5 than on the intercepts of items 1 and 3. Also, rank ordering of factor means

demonstrated that Chile had the highest factor mean of life satisfaction, while Bulgaria had the lowest. However, because of the lack of invariance on the intercepts of items 2, 4, and 5, the rank order of factor means should be interpreted with caution.

Although it was not our primary aim to investigate sources of SWLS non-invariance, we checked whether the dissimilar age mean across countries affected the intercepts of non-invariant items (items 2, 4, and 5) in a supplementary analysis. Results showed dissimilar mean age across countries significantly affected the intercept of item 4. Hultell and Gustavsson (2008) reported that item 4 (“So far I have gotten the important things I want in life”) is endorsed more by older than by younger respondents. Therefore, countries that recruit older samples may have a significantly higher intercept than countries that recruit younger samples.

We used three methods to examine measurement invariance. Although all three methods can be used for measurement invariance testing, one method may be more appropriate than another depending on sample characteristics and research questions. We suggest if researchers want to compare a small number of groups, MG-CFA works well. Because MG-CFA is commonly used and has a relatively long history as a measurement invariance technique, rules of model fit indices are established more so than with other approaches. Also, scholars (e.g., Byrne, Shavelson, & Muthen, 1989; Steenkamp & Baumgartner, 1998) have suggested that if partial scalar invariance is established with at least two invariant items in MG-CFA, factor means can be meaningfully compared (Davidov et al., 2012). However, applied researchers need to be aware of criticisms of the partial invariance approach (e.g., De Beuckelaer & Swinnen, 2011), and thorough investigation of the consequences of partial invariance is warranted.

If researchers consider the groups as a random sample of the population and are interested in generalization to the population, ML-CFA is appropriate. Furthermore, ML-CFA

allows researchers to investigate the factor structure of a measure at each level (that is, individual-level and country-level). However, model fit rules for ML-CFA have not been well-established and further research on this issue is needed. In addition, even with full or partial scalar invariance, factor means cannot be individually estimated because countries are considered as random. Although an omnibus test of factor mean equality across countries can be conducted by testing whether the between-level factor variance equals zero, specific factor mean comparisons are not possible. Lastly, if researchers want to compare the factor means of a large number of groups, even with some departure from exact measurement invariance (i.e., approximate invariance), alignment optimization is the most appropriate method because it was specifically developed for this purpose. Alignment optimization provides item-level and group-level measurement invariance information per each item beyond general model information. Again, one of the advantages of the method is that it allows researchers to compare factor means without satisfying exact scalar invariance due to the assumption of approximate measurement invariance. Thus, if the degree of non-invariance is not very severe, factor means of groups can be meaningfully compared, even though the decision criterion has not been established yet (except a rough 25% rule). In spite of the noteworthy advantages of alignment optimization, this method is still relatively new, and more validation of the 25% rule is required.

Strengths and Limitations

The current study has several strengths. By including multiple countries, we provide more generalizable results with regard to SWLS measurement invariance than reported in previous studies. We found that SWLS scalar level measurement invariance does not hold across countries. Second, three methods were used to test measurement invariance instead of a single method. With consistent results from all three methods, we were able to obtain more credible

results by ruling out the possibility that the non-invariant results were produced due to the unique features of a single method. Third, beyond the general information of measurement invariance, this study provided specific non-invariant item and non-invariant country information. By showing which items and which countries were non-invariant, an improved depth of understanding about the individual items on the life satisfaction scale was attained. Fourth, we performed partial invariance testing and reported the rank order of factor means under partial invariance.

This study has multiple strengths, but also has limitations. First, although the current study included more countries than previous studies, the data did not include countries in Africa, Central Asia, South Asia, and the Middle East. Therefore, the results may not be generalized to countries in those regions. Future researchers should replicate this study including samples from those regions to achieve a higher generalizability. Second, we were unable to examine why the intercepts of items 2, 4, and 5 were non-invariant although we demonstrated how to investigate the source of non-invariance by including potential predictors of non-invariance using the ML-MIMIC model. More research is needed to investigate sources of the non-invariant results.

Practical Implications and Suggestions

Based on our results, we provide practical suggestions and guidance. If cross-cultural researchers and practitioners use SWLS to measure and compare country means of life satisfaction, this can be done under partial invariance or approximate invariance. However, this practice should be used with caution because the impact of partial invariance on the parameters of primary interest (e.g., factor means) needs more investigation, and there is no consensus about how approximate is close enough to conclude that a measure is invariant. Because we observed some degree of non-invariance in SWLS, it is not recommended to use the observed scores of

SWLS for country mean comparisons. We showed that items 1 and 3 are invariant across the 26 countries. When researchers construct a CFA model across countries (e.g., for measurement invariance testing to explore the sources of non-invariance), these items can be chosen as an anchor or reference item. Furthermore, because metric invariance holds, researchers can use SWLS scores to compare structural relations among variables across countries (e.g., the association between life satisfaction and a certain factor is stronger in one country than another).

We recommend researchers and practitioners conduct measurement invariance testing before mean scores are compared across groups. Through measurement invariance testing, researchers and practitioners in cross-national studies could gain valuable insights on the similarity and dissimilarity across countries on their perception of a certain construct. In addition, researchers and practitioners may broaden their understanding of cross-cultural differences by focusing on locating non-invariant items and non-invariant countries and further identifying the sources of non-invariance. Such investigation would be valuable especially when cultural differences are seemingly expected and distinctive cultural factors are likely to affect item responses (e.g., modesty and self-enhancement; Marsh, Hau, Artelt, Baumert, & Peschar, 2006). Finally, understanding of non-invariance will guide researchers and practitioners to develop more culturally-invariant items of scales in the future.

Conclusion

We investigated the Satisfaction with Life Scale across 26 countries using three different measurement invariance testing methods. Results consistently revealed that scalar measurement invariance did not hold across countries. With partial invariance testing, non-invariant intercepts were identified and factor means were compared under partial invariance. Results of the current study can be used to help guide future studies of measurement invariance across countries.

References

- Andrews F. M. & Whitney S. B. (1976). *Social Indicators of Well-Being: Americans' Perception of Life Quality*. Plenum Press, New York, NY
- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495-508.

- Byrne, B. M. (1994). Testing for the factorial validity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslach Burnout Inventory. *Multivariate Behavioral Research, 29*, 289–311.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.
- Dawes, J. G. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International journal of market research, 51*(1), 61-77.
- De Beuckelaer, A., & Swinnen, G. (2011). Biased latent variable mean comparisons due to measurement noninvariance: A simulation study. *European Association for Methodology Series, 117-147*.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542-575.
- Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71-75.
- Douglas, S. P., & Craig, C. S. (1983). Examining performance of US multinationals in foreign markets. *Journal of International Business Studies, 14*(3), 51-62.
- Eid, M., Langeheine, R., & Diener, E. (2003). Comparing Typological Structures Across Cultures By Multigroup Latent Class Analysis A Primer. *Journal of Cross-Cultural*

- Psychology*, 34(2), 195-210.
- Fredrickson, B. L., Tugade, M. M., Waugh, C. E., & Larkin, G. R. (2003). What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *Journal of Personality and Social Psychology*, 84(2), 365-376.
- Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, 106(4), 766-794.
- Hofer, J., Chasiotis, A., & Campos, D. (2006). Congruence between social values and implicit motives: Effects on life satisfaction across three cultures. *European Journal of Personality*, 20(4), 305-324.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117-144.
- Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist*, 1, 179-188.
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage Publications.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Hultell, D., & Gustavsson, J. P. (2008). A psychometric evaluation of the Satisfaction with Life Scale in a Swedish nationwide sample of university students. *Personality and Individual Differences*, 44(5), 1070-1079.

- Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(2), 265-282.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224-253.
- Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, 6(4), 311-360.
- Marsh, H. W., Vallerand, R. J., Lafrenière, M. A. K., Parker, P., Morin, A. J., Carbonneau, N., ... & Salah Abduljabbar, A. (2013). Passion: Does one scale fit all? Construct validity of two-factor passion scale and psychometric invariance over different activities and languages. *Psychological Assessment*, 25(3), 796-809.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Meredith, W., & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, 57(2), 289-311.
- Millsap R. E. (2011). *Statistical Approaches to Measurement Invariance*. New York, NY: Routledge.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22(3), 376-398.
- Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology*, 5, 978. doi: 10.3389/fpsyg.2014.00978

- Oishi, S. (2006). The concept of life satisfaction across cultures: An IRT analysis. *Journal of Research in Personality, 40*(4), 411-423.
- Oishi, S., Diener, E. F., Lucas, R. E., & Suh, E. M. (1999). Cross-cultural variations in predictors of life satisfaction: Perspectives from needs and values. *Personality and Social Psychology Bulletin, 25*(8), 980-990.
- Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment, 5*(2), 164-172.
- Pavot, W., Diener, E. D., Colvin, C. R., & Sandvik, E. (1991). Further validation of the satisfaction with life scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment, 57*(1), 149-161.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological Bulletin, 114*(3), 552-566.
- Roster, C., Albaum, G., & Rogers, R. (2006). Can cross-national/cultural studies presume etic equivalency in respondents' use of extreme categories of Likert rating scales? *International Journal of Market Research, 48*(6), 741-759.
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*(1), 31-57.
- Shin, D. C., & Johnson, D. M. (1978). Avowed happiness as an overall assessment of the quality of life. *Social Indicators Research, 5*(1-4), 475-492.
- Spector, P. E., Allen, T. D., Poelmans, S. A. Y., Lapierre, L. M., Cooper, C. L., O'Driscoll, M., ... & Widarszal-Bazyl, M. (2007). Cross-national differences in relationships of work

- demands, job satisfaction and turnover intentions with work-family conflict. *Personnel Psychology*, 60(4), 805-835.
- Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78-90.
- Van de Vijver, F. J., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage Publication.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.
- Yoon, M., & Kim, E. S. (2014). A comparison of sequential and nonsequential specification searches in testing factorial invariance. *Behavior Research Methods*, 46(4), 1199-1206.

Table 1. *Factor Mean Comparisons in Partial Scalar MG-CFA, Alignment Optimization, and Scale Composite Scores*

Method		Partial Scalar MG-CFA		Alignment Optimization		Scale Composite Scores	
Ranking	Country	Factor Mean	Country	Factor Mean	Country	Mean	
1	Chile	1.07	Chile	.81	Chile	5.54	
2	Netherland	.86	Netherland	.62	Netherland	5.27	
3	Finland	.73	Puerto Rico	.48	Puerto Rico	5.13	
4	Puerto Rico	.70	Finland	.47	Finland	5.09	
5	Peru	.62	Greece	.39	Greece	5.02	
6	Greece	.58	Peru	.37	Peru	5.01	
7	New Zealand	.58	New Zealand	.37	New Zealand	4.98	
8	Turkey	.58	Argentina	.35	Argentina	4.97	
9	Argentina	.52	Canada	.27	Canada	4.85	
10	Taiwan	.46	Estonia	.26	Estonia	4.85	
11	Canada	.44	Bolivia	.22	Bolivia	4.77	
12	Bolivia	.41	Turkey	.22	Australia	4.76	
13	Estonia	.33	Australia	.18	Turkey	4.73	
14	Slovenia	.28	Poland	.06	Romania	4.58	
15	Poland	.26	Taiwan	.06	US	4.50	
16	Australia	.25	Romania	.03	Slovenia	4.48	
17	Romania	.04	Slovenia	.00	Poland	4.45	
18	US	.00	US	.00	Taiwan	4.43	
19	UK	-.31	UK	-.12	UK	4.36	
20	Hong Kong	-.45	Hong Kong	-.41	Hong Kong	3.89	
21	Korea	-.58	Korea	-.44	Korea	3.85	
22	China	-.59	China	-.53	Ukraine	3.76	
23	Ukraine	-.72	Japan	-.59	China	3.63	
24	Japan	-.76	Ukraine	-.61	Japan	3.58	
25	Bulgaria	-.91	Spain	-.74	Bulgaria	3.58	
26	Spain	-.99	Bulgaria	-.78	Spain	3.56	

Table 2. *Alignment Fit Statistics*

Item	Intercepts		Loadings	
	Fit function contribution	R-Square	Fit function contribution	R-Square
FIXED approach				
Life satisfaction- item 1	-142.333	0.92	-128.812	0.80
Life satisfaction- item 2	-144.959	0.92	-111.414	0.94
Life satisfaction- item 3	-127.450	0.97	-129.955	0.70
Life satisfaction- item 4	-180.664	0.86	-144.641	0.48
Life satisfaction- item 5	-194.023	0.74	-136.034	0.56

Table 3. *Approximate Measurement Invariance (Non-invariance) for Groups*

Item	Invariance (Non-invariance) for Groups
FIXED approach (including 26 countries)	
Intercepts	
LS1	AR AU BO BG CA CL CN (EE) FI GR HK JP KR (NL) NZ PE PL PR RO SI ES (TW) TR GB US UA
LS2	AR (AU) BO BG (CA) CL CN (EE) (FI) GR HK (JP) (KR) (NL) (NZ) PE PL PR RO SI ES (TW) TR (GB) US UA
LS3	AR (AU) BO BG CA CL CN EE FI GR HK JP KR NL NZ PE PL PR (RO) (SI) ES TW TR GB US UA
LS4	(AR) (AU) BO BG CA (CL) CN (EE) FI (GR) HK (JP) KR NL NZ PE PL PR (RO) SI ES (TW) TR (GB) (US) UA
LS5	AR AU BO (BG) CA CL (CN) EE FI GR HK (JP) (KR) NL NZ PE PL PR RO SI (ES) (TW) TR GB US (UA)
Loadings	
LS1	AR AU BO (BG) CA CL CN EE FI GR HK JP KR NL NZ PE PL PR RO SI ES TW TR GB US UA
LS2	AR AU BO BG CA CL CN EE FI GR HK JP KR NL NZ PE PL PR RO SI ES TW TR GB US UA
LS3	AR AU BO BG CA (CL) CN EE FI GR HK JP KR (NL) NZ PE PL PR RO (SI) ES TW (TR) GB US UA
LS4	AR AU BO BG CA CL (CN) EE FI (GR) HK JP KR NL NZ PE PL PR RO SI ES TW TR GB US UA
LS5	AR AU BO BG CA CL CN EE FI GR HK (JP) KR NL NZ PE PL PR RO SI ES TW TR GB US UA

Note. LS = life satisfaction, AR=Argentina, AU=Australia, BO=Bolivia, BG=Bulgaria, CA=Canada, CL=Chile, CN=China, EE=Estonia, FI=Finland, GR=Greece, HK=Hong Kong, JP=Japan, KR=South Korea, NL=Netherland, NZ=New Zealand PE=Peru, PL=Poland, PR=Puerto Rica, RO=Romania, SI=Slovenia, ES=Spain, TW=Taiwan, TR=Turkey, GB=United Kingdom, US=United States, UA=Ukraine

Online Supplement A
Demographic Information for 26 Countries

Table A: *Demographic Information for 26 Countries*

Country	Total (N)	Males (%)	Males (SD)	Age (Mean)	Age (SD)	Cronbach's Alpha
Argentina	356	79%	.41	42.27	8.37	0.89
Australia	502	50%	.50	39.78	11.76	0.90
Bolivia	245	44%	.50	32.88	9.84	0.86
Bulgaria	286	65%	.48	38.95	8.62	0.60
Canada	214	58%	.49	44.02	8.38	0.89
Chile	191	73%	.45	43.5	8.42	0.89
China	251	57%	.50	33.41	6.92	0.82
Estonia	249	44%	.50	35.67	6.42	0.86
Finland	259	46%	.50	45.85	8.27	0.88
Greece	249	43%	.50	38.71	8.92	0.89
Hong Kong	200	47%	.50	30.49	11.04	0.84
Japan	241	68%	.47	40.95	9.48	0.88
Korea	256	61%	.49	33.82	8.94	0.89
Netherland	257	79%	.41	43.55	9.78	0.85
New Zealand	468	69%	.46	49.03	10.61	0.90
Peru	171	76%	.43	42.01	8.94	0.80
Poland	261	52%	.50	39.97	10.61	0.86
Puerto Rica	250	43%	.50	41.88	10.98	0.90
Romania	271	56%	.50	35.61	10.77	0.88
Slovenia	306	94%	.23	39.51	7.58	0.89
Spain	458	77%	.42	42.82	7.64	0.87
Taiwan	265	39%	.49	32.05	9.90	0.91
Turkey	262	71%	.45	40.1	8.91	0.87
United Kingdom	137	69%	.46	39.28	6.58	0.86
United States	171	45%	.50	41.91	10.37	0.90
Ukraine	228	43%	.50	35.86	10.99	0.88
Mean	268.04	61%	.49	39.80	10.44	0.86

Note. $n = 7004$; SD = standard deviation

Online Supplement B
A Syntax For MG-CFA

TITLE: MG-CFA

DATA: FILE IS CISMS.txt;

VARIABLE: NAMES = Coun LS1 LS2 LS3 LS4 LS5 Age;

USEVARIABLES ARE LS1 LS2 LS3 LS4 LS5;

Grouping is Coun(0=US 1=Argentina 2=Australia 3=Bolivia 4=Bulgaria 5=Canada 6=Chile
7=China 8=Estonia 9=Finland 10=Greece 11=Hong Kong 12=Japan 13=Korea 14=Netherlan
15=NewZeal 16=Peru 17=Poland 18=PuertoRi 19=Romania 20=Slovenia 21=Spain
22=Taiwan 23=Turkey 24=UK 26=Ukraine);

MISSING = ALL (999);

ANALYSIS:

ESTIMATOR = ML;

MODEL= configural metric scalar;

MODEL:

F1 by LS1 LS2 LS3 LS4 LS5;

Online Supplement C
A Syntax For ML-CFA

```
TITLE: ML-CFA configural
DATA: FILE IS CISMS.txt;
VARIABLE: NAMES = Coun LS1 LS2 LS3 LS4 LS5 Age;
USEVARIABLES ARE LS1 LS2 LS3 LS4 LS5;
MISSING = ALL (999);
Cluster = Coun;
ANALYSIS: type = twolevel;
MODEL:
% within%
FW by LS1 LS2 LS3 LS4 LS5;
% between%
FB by LS1 LS2 LS3 LS4 LS5;
OUTPUT:SAMPSTAT RESIDUAL STDYX modindices(all,0) TECH1;
```

```
TITLE: ML-CFA metric
DATA: FILE IS CISMS.txt;
VARIABLE: NAMES = Coun LS1 LS2 LS3 LS4 LS5 Age;
USEVARIABLES ARE LS1 LS2 LS3 LS4 LS5;
MISSING = ALL (999);
Cluster = Coun;
ANALYSIS: type = twolevel;
MODEL:
% within%
FW by LS1 LS2(1) LS3(2) LS4(3) LS5(4);
% between%
FB by LS1 LS2(1) LS3(2) LS4(3) LS5(4);
OUTPUT:SAMPSTAT RESIDUAL STDYX modindices(all,0) TECH1;
```

```
TITLE: ML-CFA scalar
DATA: FILE IS CISMS.txt;
VARIABLE: NAMES = Coun LS1 LS2 LS3 LS4 LS5 Age;
USEVARIABLES ARE LS1 LS2 LS3 LS4 LS5;
MISSING = ALL (999);
Cluster = Coun;
ANALYSIS: type = twolevel;
MODEL:
% within%
FW by LS1 LS2(1) LS3(2) LS4(3) LS5(4);
% between%
FB by LS1 LS2(1) LS3(2) LS4(3) LS5(4);
LS1@0; LS2@0; LS3@0; LS4@0; LS5@0;
OUTPUT:SAMPSTAT RESIDUAL STDYX modindices(all,0) TECH1;
```

Online Supplement D
A Syntax For Alignment Optimization

```
TITLE: Alignment optimization
DATA: FILE IS CISMS.txt;
VARIABLE: NAMES = Coun LS1 LS2 LS3 LS4 LS5 Age;
USEVARIABLES ARE LS1 LS2 LS3 LS4 LS5;
MISSING = ALL (999);
CLASSES = c(26);
KNOWNCLASS = c(coun);
ANALYSIS: TYPE = MIXTURE;
ESTIMATOR = ML;
ALIGNMENT = fixed(0);
ASTARTS = 100
MODEL: %OVERALL%
F1 by LS1 LS2 LS3 LS4 LS5;
OUTPUT: TECH1 TECH8 ALIGN;
```