CrossMark

RESEARCH PAPER

Measurement Invariance of Personal Well-Being Index (PWI-8) Across 26 Countries

Magdalena Żemojtel-Piotrowska[1] · Jarosław P. Piotrowski[2] ·
Jan Cieciuch[3,4] · Byron G. Adams[5,6] · Evgeny N. Osin[7] ·
Rahkman Ardi[8] · Sergiu Bălţătescu[9] · Arbinda Lal Bhomi[10] ·
Amanda Clinton[11] · Gisela T. de Clunie[12] · Carla Esteves[13] ·
Valdiney Gouveia[14] · Ashraf Hosseini[15] · Hooria Seyedhosseini Ghaheh[16] ·
Narine Khachatryan[17] · Shanmukh Vasant Kamble[18] ·
Anna Kawula[19] · Kadi Liik[20] · Eva Letovancova[21] ·
Sara Malo Cerrato[22] · Carles Alsinet Mora[23] · Sofya Nartova-Bochaver[7] ·
Marija Nikolic[24] · Joonha Park[25] · Elena Paspalanova[26] ·
Győző Pék[27] · Joanna Różycka-Tran[1] · Ha Truong Thi Khanh[28] ·
Takashi Tsubakita[25] · Christin-Melanie Vauclair[13] ·
Anna Włodarczyk[29,30] · John Maltby[31]

Abstract This report examines the measurement invariance of the Personal Well-being Index with 8 items (PWI-8). University students (N = 5731) from 26 countries completed the measure either through paper and pencil or electronic mode. We examined uni-dimensional structure of PWI and performed a Multi-group CFA to assess the measurement invariance across the 26 countries, using conventional approach and the alignment

---

Order of authors reflects their contribution to the work.

✉ Magdalena Żemojtel-Piotrowska
    psymzp@ug.edu.pl

[1] University of Gdansk, Bazynskiego 4, 80-952 Gdańsk, Poland

[2] Poznan Faculty, University of Social Sciences and Humanities, Poznan, Poland

[3] Cardinal Stefan Wyszyński University in Warsaw, Warsaw, Poland

[4] Zurich University, Zurich, Switzerland

[5] Tilburg University, Tilburg, The Netherlands

[6] University of Johannesburg, Johannesburg, South Africa

[7] National Research University Higher School of Economics, Moscow, Russia

[8] Airlangga University, Surabaya, Indonesia

[9] University of Oradea, Oradea, Romania

procedure. The findings provide evidence of configural and partial metric invariance, as well as partial scalar invariance across samples. The findings suggest that PWI-8 can be used to examine correlates of life satisfaction across all included countries, however it is impossible to compare raw scores across countries.

Keywords Personal well-being index · Cross-cultural studies · Measurement invariance

1 Introduction

Quality of life has become an important, well-researched topic over the last few years. More specifically this has been examined in terms of well-being, which is often assessed at national levels for international comparisons. These international comparisons, however, require measures which have been shown to be invariant across different cultural groups and countries. The objective of this study is to examine the measurement invariance of just such a measure, the Personal well-being index (PWI), which is considered one of the most popular measures for evaluating subjective well-being (International Well-being Group 2013; Sirgy 2012).

Life satisfaction is the cognitive component of subjective well-being and has a general character (Diener 1984). According to Diener et al. (1985), life satisfaction is the effect of a judgmental process in which "a comparison of one's circumstances with what is thought to be appropriate standard" (p. 71) is made. Therefore, it refers to some standards of evaluation, which could be related to different life domains. PWI is the decomposition of life satisfaction in satisfactions with different domains (Cummins et al. 2003), namely: (1)

[10]   Tribhuvan University, Kathmandu, Nepal

[11]   University of Puerto Rico, Puerto Rico, PR, USA

[12]   Universidad Tecnológica de Panamá, Panamá, Panama

[13]   Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL, Lisboa, Portugal

[14]   Federal University of Paraíba, João Pessoa, Brazil

[15]   University of Melbourne, Parkville, Australia

[16]   Islamic Azad University, Tehran, Iran

[17]   Yerevan State University, Yerevan, Armenia

[18]   Karnatak University, Dharwad, India

[19]   Pedagogic University in Cracov, Kraków, Poland

[20]   Talinn University, Talinn, Estonia

[21]   Comenius University in Bratislava, Bratislava, Slovakia

[22]   University of Girona, Girona, Spain

[23]   University of Lleida, Lleida, Spain

[24]   LUM University, Casamassima, Italy

[25]   Nagoya University of Commerce and Business, Nagoya, Japan

[26]   New Bulgarian University, Sofia, Bulgaria

standard of living, (2) personal health, (3) life achievements, (4) personal relationships, (5) personal safety, (6) community connectedness, (7) future security, and (8) religion and spirituality. The PWI has been used as an assessment of life satisfaction in child (Casas et al. 2012), adolescent and student (Tomyn et al. 2011), aging (Bricker-Katz et al. 2009; Forjaz et al. 2011), and clinical populations (e.g. Engel and Cummins 2011; Werner 2012). The scale is intended to be inclusive of all important life domains which could contribute to the general level of life satisfaction and to serve as a tool in cross-cultural comparisons on relative importance of particular domains in life satisfaction (International Wellbeing Group 2013). The idea behind developing the PWI was to include the most important predictors of general life satisfaction. A selection of domains was done by an international team and was based on several criteria: the selection should include only basic domains important for predicting "life satisfaction as a whole"; each domain should refer to broad aspects of life, and each domain needs to represent an indicator and not a causal variable of general life satisfaction (see International Wellbeing Group 2013).

1.1 Measurement Invariance as Means of Cross-Cultural Inquiry

Oishi (2010) pointed towards several important methodological and conceptual issues related to cross-cultural studies on subjective well-being; these are conceptual equivalence, translation issues, desirability of the concept, response style, item functioning, differences in self-presentations, memory bias, and validity criteria. For instance, single-item measures of subjective well-being like Cantrill's ladder or general items on life satisfaction are less reliable than longer scales and do not allow for more in-depth examination of cross-cultural equivalence in terms of measurement equivalence (see Oishi 2010). As the PWI is a multi-item scale it is particularly useful in cross-cultural research. Based on the work of an international group of well-being researchers, it uses items that are simple and easy to translate, which allows for minimization of problems with conceptual equivalence and translations (see International Wellbeing Group 2006, 2013).

Multi-group Confirmatory Factor Analysis (MGCFA) serves as a statistical tool for assessment of cross-cultural equivalence of a measure. Such analysis is fundamental for establishing the usefulness of any measure intended for cross-cultural research. There are three levels of measurement invariance which are most commonly used to establish whether a measure is equivalent: (a) Configural invariance provides indication that the general factor structure of the measure is the same across different groups. At this level, the construct is measured similarly in different samples. (b) Metric invariance indicates that the factor loading of items is similar, (i.e., load in the same way in assumed factor) across groups. At this level, measure correlates and/or predictors may be compared across samples. (c) Scalar invariance indicates that item intercepts are equal intercepts across groups. At this level means may be compared across samples (Davidov et al. 2014).

Scalar invariance is rarely found in large cross-cultural comparisons (see Davidov et al. 2014). In general, there are very few studies examining the measurement invariance of

[27]   University of Debrecen, Debrecen, Hungary

[28]   Vietnam National University, Hanoi, Vietnam

[29]   University of Basque Country, Leioa, Spain

[30]   University of Santiago de Chile, Santiago, Chile

[31]   University of Leicester, Leicester, England

scales measuring subjective well-being. For instance, only metric invariance was reported for Satisfaction with Life Scale (SWLS) across samples from US, England, and Japan (Whisman and Judd 2016), and mixed results for SWLS invariance across Russian and US samples (Tucker et al. 2006). Most studies typically focus on comparisons between several national groups, rarely examining large representation of countries (see also Ponizovsky et al. 2013).

1.2 The Current Study

Despite the many cultural adaptations of this measure (see International Wellbeing Group 2013, for details) and its increasing popularity among cross-cultural researchers, to the best of our knowledge, there is little evidence that the PWI is invariant across different countries, with the exceptions of adolescent samples in Chile and Brazil (Sarriera et al. 2014) or general populations in Hong Kong and Australia (Lau et al. 2005). Sometimes the levels of PWI are compared without examination of measurement invariance, as in the case of Romania and Hungary (Baltatescu 2014). This report intends to fill this gap by examining the measurement invariance of the PWI in university student samples across 26 countries.

In the current study, we examine the measurement invariance of PWI across countries from different regions of the world: Europe (10), Asia (10), Africa (2), and Latin America (4). Among them, there are the most affluent and developed countries, like UK or Japan, and less affluent, agrarian societies like Iran or Kenya. In terms of cultural regions we had representatives for all Huntington (1996) cultural groups (i.e., Western, Orthodox, Confucian, Japanese, Latin American, Hindu, Buddhists, Islamic, African, and Sinic), and in terms of religion we had countries representing all main world religions. This selection of countries is not exhaustive, but it allows for examining measurement invariance of PWI across different languages and cultures. The aim of the study was to investigate the measurement invariance of the PWI across different countries and languages. Given the large number of countries compared, we expected to find support only for the metric level of invariance, as scalar invariance is hardly found in large cross-national comparisons.

2 Method

2.1 Sample and Procedure

Data were collected in a paper–pencil or online formats between April 2014 and August 2015. The sample comprised 5,530 university students (42.4 % men, $M = 21.29$, $SD = 3.15$, age ranged from 16 to 39). We excluded all participants above the age of 40 (1.7 % of total sample) from the analyses, as in most countries the respondents' age was in the 18–25 range, and rarely exceeded 30 years. We also asked students to indicate the socioeconomic status of their families on a 7-point Likert-type scale (from 1 = *significantly below average* to 7 = *significantly over average*). The students majored in different fields (e.g., social sciences, technical sciences, and medical sciences) and originated from 26 countries (see Table 1 for a sample breakdown). They were recruited for the study during their classes and participated on a voluntary basis. They completed the PWI as part of a broader project on entitlement and subjective well-being. The paper-and-pencil surveys were administered in small groups ($n < 15$). In countries where the survey was

Table 1 Sample information and internal consistency coefficients for PWI-8 in 26 countries

Country	N	Female (%)	Age M	SES	Language	Procedure	Cronbach's α
Armenia	223	48	19.00	4.98	Armenian	Paper–pencil	.84
Brazil	225	64	20.50	4.38	Portuguese	Online	.82
Bulgaria	197	66	23.70	4.66	Bulgarian	Paper–pencil	.82
Chile	241	52	22.00	4.34	Spanish	Paper–pencil	.82
Estonia	289	68	22.22	4.37	Estonian	Online	.79
Hungary	206	69	21.01	n/a	Hungarian	Paper–pencil	.81
India	200	69	22.59	4.37	English	Paper–pencil	.76
Indonesia	200	50	21.38	4.70	Indonesian	Online	.89
Iran	201	50	21.28	4.46	English	Paper–pencil	.75
Israel	200	40	24.58	4.88	Hebrew	Online	.82
Japan	202	23	18.91	4.14	Japanese	Paper–pencil	.82
Kenya	161	53	23.39	4.07	English	Paper–pencil	.90
Nepal	199	51	22.70	4.08	English	Paper–pencil	.74
Panama	176	32	22.03	4.13	Spanish	Online	.88
Poland	258	60	21.85	4.69	Polish	Paper–pencil	.86
Portugal	187	77	22.79	4.11	Portuguese	Online	.82
Puerto Rico	300	43	20.26	4.16	Spanish	Paper–pencil	.84
Romania	210	48	21.49	4.72	Romanian	Paper–pencil	.85
Russia	227	83	21.03	3.11	Russian	Online	.87
Serbia	199	61	22.46	3.77	Serbian	Paper–pencil	.84
Slovakia	202	72	21.13	4.76	Slovakian	Paper–pencil	.77
S. Africa	188	67	20.17	4.45	English	Paper–pencil	.79
S. Korea	215	55	22.20	3.90	Korean	Paper–pencil	.83
Spain	196	51	21.20	4.01	Spanish (Catalan)	Online	.72
UK	302	81	19.44	4.21	English	Online	.84
Vietnam	259	53	20.52	4.25	Vietnamese	Paper–pencil	.86

Due to technical error data for socio-economic status of family are not available for Hungary

SES subjective economic status of family (1–7)

administered in a non-native language, a researcher assisted students and explained the meaning of particular words. This was the case of India, Iran, Kenya, Nepal, and South Africa; however, only in Iran and Nepal is English not an official language.

2.2 Measure

The personal well-being index (PWI; Cummins et al. 2003; International Wellbeing Group 2013) measures satisfaction with different life domains: (1) standard of living, (2) health, (3) life achievements, (4) personal relationships, (5) personal safety, (6) community connectedness, (7) future security, and (8) religion and spirituality. Previous studies suggest that in different countries the relative importance of religiosity and spirituality varies as a function of cultural differences (Norris and Inglehart 2004) and that they both significantly contribute to subjective well-being (Casas et al. 2009; Piedmont and Friedman

2012). Therefore, we used one combined question about religiosity and spirituality (*How much are you satisfied with your spirituality or religion?*), as suggested by manual for PWI-8, despite some researchers' postulating two parallel versions for item 8 (Sarriera et al. 2014). Participants responded on an 11-point Likert-type scale (0 = *not at all satisfied* to 10 = *totally satisfied*). National versions of scale were authorised versions or they were obtained by repeating back translation procedures with bilingual researchers and with the participation of Robert Cummins (see Table 1 for information on language of administration).

2.3 Statistical Analyses

We started by using Mplus 7.4 to perform Confirmatory Factor Analysis (CFA) in order to test for a unidimensional structure of the PWI-8 in each country sample. Because the score distributions were not perfectly normal and Mardia's multivariate skewness and kurtosis statistics (presented in Supplementary Information, table SI.3) were significant in all samples, we used the robust Satorra-Bentler χ^2 (Satorra and Bentler 1994; referred to as estimator MLM in Mplus). Because the country samples differed in gender distribution, we used weighting in all analyses to equalize the contribution of male and female respondents within each country to the model (the weights were calculated to achieve a target $N = 100$ for males and females in each group).

The model fit was examined using the most common fit indices: the Chi square (χ^2), the CFI (Comparative Fit Index), the RMSEA (root mean squared error of approximation), and the SRMR (standardized root mean squared residual). In larger samples (N > 200), practical fit indices (CFI, RMSEA, and SRMR) are preferred to the χ^2 as they are less sensitive to sample size (Chen, 2007; Davidov et al. 2014). CFI values above .90 were considered as evidence of an acceptable model fit and those above .95 as evidence of a good fit. Because in smaller samples RMSEA tends to over-reject correct models (Hu and Bentler 1999), we used RMSEA values of .10 and .06 as thresholds for acceptable and good fit, respectively. For SRMR, .08 and .05 thresholds were used (Brown 2015). In samples where the fit of the one-factor theoretical model was outside the acceptable range and a pronounced and interpretable outlier was found among the modification indices, suggesting an error covariance, the latter was added and the model was retested. We aimed to introduce as few modifications as possible in order to achieve acceptable fit without over-complicating the model.

After establishing the measurement model for each country, we proceeded by conducting multi-group CFA (MGCFA) to test for configural, metric, and scalar measurement invariance. We tested the invariance based on modified measurement models using a conventional approach (Byrne 2012). We used the ΔCFI and ΔRMSEA values of .010 and .015, respectively, as evidence of pronounced difference between nested models (Chen 2007; Cheung and Rensvold 2002). We looked for outliers among the modification indices and introduced them into the model one-by-one, until the difference in practical fit indices between the configural invariance and partial metric invariance models became small enough (ΔCFI ≤ .01, ΔRMSEA ≤ .015). The procedure was repeated for scalar invariance.

A potential drawback of the manual approach is that each modification results in a different model, and the exact resulting list of non-invariant parameters is dependent on the sequence in which modifications are entered into the model. In case of a long sequence of modifications, the conventional approach (addressing the strongest modification index at each step) does not guarantee that the resulting model will be optimal (i.e., simplest, with

the fewest number of non-invariant parameters). This problem is overcome by the alignment procedure (Asparouhov and Muthén 2014), which evaluates different combinations of non-equivalent parameters to find an optimal model. We tried to cross-validate our findings using the alignment procedure, based on the same modified measurement model. Finally, we tested the invariance of the PWI across genders. Because the sample sizes were not large enough to test the invariance across genders in each country separately, we tested a single-factor model in the combined sample with robust Chi square (MLR) and standard errors computed using the sandwich estimator for clustered samples to account for non-independence of observations within countries.

3 Results

3.1 Scale Structure Across Countries

The internal consistency (Cronbach's alpha) values of the PWI in each national sample are presented in Table 1. Cronbach's alpha values above .70, indicating good reliability (Lance et al. 2006), were found in all samples. To ensure unidimensionality, we conducted parallel analysis. In all 26 samples, only the first eigenvalue exceeded the one obtained for random data using parallel analysis, and all the items exhibited significant loadings on the single dimension.

Table 2 presents the results of single-sample CFA analyses for the initial (theoretical) model. In most countries, the theoretical model showed acceptable fit, based on the combination of practical fit indices. The fit of the model was outside the acceptable range in Spain, Poland, South Korea, Hungary, Romania, Indonesia, and Panama.

We explored the modification indices in countries with unacceptable and marginal fit and introduced additional covariances in cases where they were theoretically justified and a strong ($\Delta\chi^2 > 10$) outlier was found among modification indices. The error covariance for items four and five (relationships and safety) was found in three Hispanic countries, in line with previous studies (Sarriera et al. 2014). The error covariance of items four and six (relationships and feeling part of community) was peculiar to two post-Communist Central European countries (Poland, Hungary). The other error covariances were explained by back-translation analysis. For instance, the error covariance of items five and seven was found in countries (Poland, Brazil) where local translations used the same word for "safety" and "security". In South Korea, items one, three, and seven, reflecting satisfaction with financial success, were associated. We added two error covariances to address this weak subdimension. The introduction of additional error covariances resulted in acceptable fit in all countries (shown in Table 3).

3.2 Measurement Invariance Analyses Across Countries

We proceeded by conducting invariance analyses. The multi-group model included modified measurement models for 10 countries and theoretical model for the remaining 16 countries. The configural model showed good fit to the data, the fit of the metric model was acceptable, and the fit of the scalar model was poor (see Table 4). The Chi square differences were also significant ($p < .001$) between the three models. The difference in practical fit indices between the configural and metric models was very small for RMSEA (ΔRMSEA = .003), but above the recommended .01 threshold for the CFI (ΔCFI = .018)

Table 2 Goodness-of-fit indices for PWI8 one-factor model (no error covariances)

Country	S-B χ^2	SCF	CFI	RMSEA (90 % CI)	SRMR
Armenia	59.03***	1.67	.901	.095 (.068–.124)	.056
Brazil	51.72***	1.35	.912	.089 (.059–.119)	.052
Bulgaria	47.66**	1.24	.924	.084 (.053–.115)	.050
Chile	34.81*	1.39	.966	.056 (.022–.086)	.036
Estonia	55.76***	1.35	.902	.079 (.055–.104)	.054
Hungary	53.97***	1.47	.883	.091 (.062–.121)	.061
India	23.17	1.65	.981	.028 (.000–.070)	.044
Indonesia	60.52***	1.35	.931	.101 (.072–.130)	.047
Iran	33.88*	1.15	.953	.059 (.020–.092)	.045
Israel	37.49*	1.57	.931	.067 (.032–.100)	.050
Japan	26.22	1.73	.975	.042 (.000–.082)	.046
Kenya	28.99	3.14	.962	.055 (.000–.096)	.053
Nepal	35.05*	1.47	.916	.064 (.025–.098)	.055
Panama	55.70***	1.51	.932	.101 (.070–.133)	.045
Poland	65.24***	1.40	.904	.103 (.076–.131)	.053
Portugal	32.57*	1.62	.949	.058 (.014–.094)	.050
Puerto Rico	71.21***	1.69	.908	.092 (.070–.116)	.051
Romania	48.18***	2.17	.898	.084 (.054–.115)	.060
Russia	40.46**	1.85	.935	.077 (.042–.111)	.047
Serbia	47.56**	1.38	.927	.084 (.053–.115)	.051
Slovakia	33.80*	1.49	.939	.059 (.020–.092)	.051
South Africa	37.82**	1.57	.917	.069 (.034–.103)	.056
South Korea	99.00***	1.35	.844	.137 (.111–.164)	.064
Spain	86.24***	1.30	.756	.131 (.104–.160)	.076
UK	33.69*	1.71	.973	.048 (.016–.075)	.033
Vietnam	44.88**	1.37	.956	.070 (.042–.097)	.041

S-B χ^2 Satorra-Bentler Chi square ($df = 20$), *SCF* scaling correction factor for Satorra-Bentler χ^2, *CFI* comparative fit index, *RMSEA* root mean square of approximation, *SRMR* standardized root mean square residual

*** $p < .001$; ** $p < .01$; * $p < .05$

and we followed by establishing partial metric invariance. After six constraints for non-invariant loadings (listed in supplementary material) were relaxed, the difference in practical fit indices between the configural invariance and partial metric invariance models became small enough (ΔCFI = .01). We followed by establishing the partial scalar invariance. After relaxing 74 constraints for equal intercepts, the corresponding modification indices became non-significant at $p < .01$ level and we stopped the procedure to reduce the risk of false positives. Even though the ΔCFI criterion was not reached (ΔCFI = .024), the ΔRMSEA and ΔSRMR were quite small (<.010), and practical fit indices (Table 4) were within acceptable limits. Because these indices are to be interpreted in combination (Brown 2015), we deemed the resulting partial scalar invariance model acceptable. The complete list of non-invariant parameters obtained from the final partial scalar invariance model is given in Supplementary Information (Table SI.1). The item

Table 3 Goodness-of-fit Indices for PWI8 one-factor solution, modified measurement models

Country	S-B χ^2	df	SCF	CFI	RMSEA (90 % CI)	SRMR	Error covariances
Armenia	25.81	19	1.62	.983	.041 (.000–.077)	.041	6–8
Brazil	36.01*	19	1.27	.953	.067 (.032–.099)	.043	5–7
Hungary	38.73**	19	1.45	.932	.071 (.038–.103)	.055	4–6
Indonesia	40.52**	19	1.36	.963	.075 (.043–.107)	.041	3–7
Panama	46.32***	19	1.55	.948	.090 (.058–.124)	.043	4–5
Poland	33.86*	18	1.36	.966	.064 (.029–.097)	.039	5–7, 4–6
Puerto Rico	51.12***	19	1.69	.942	.075 (.051–.100)	.043	4–5
Romania	35.51*	19	2.18	.940	.066 (.030–.099)	.047	1–2
South Korea	54.32***	18	1.35	.928	.098 (.069–.128)	.048	1–3, 3–7
Spain	39.24**	19	1.33	.925	.074 (.041–.108)	.052	4–5

S-B χ^2 Satorra-Bentler robust Chi square (df = 20), *SCF* scaling correction factor for Satorra-Bentler χ^2, *CFI* comparative fit index, *RMSEA* root mean square of approximation, *SRMR* standardized root mean residual

*** $p < .001$; ** $p < .01$; * $p < .05$

Table 4 Fit indices for multi-group confirmatory factor analysis (MGCFA) models of the PWI8 across 26 countries

Model	χ^2	df	SCF	CFI	RMSEA (90 % CI)	SRMR
Configural	975.95	508	1.57	.948	.066 (.060–.073)	.047
Metric	1316.10	683	1.53	.930	.067 (.061–.072)	.084
Partial metric	1237.03	677	1.52	.938	.063 (.057–.068)	.075
Scalar	3455.53	852	1.22	.711	.121 (.111–.125)	.120
Partial scalar	1554.46	777	1.34	.914	.069 (.064–.074)	.079

Difference tests	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
Metric versus configural	340.47	175	.018	.001	.037
Partial metric versus configural	254.10	169	.010	.003	.028
Scalar versus partial metric	39,298.62	175	.227	.058	.045
Partial scalar versus partial metric	1669.61	100	.024	.006	.004

The list of error covariances included in the model is given in Table 2. The list of non-invariant loadings and intercepts is given in Table SI.1

SCF scaling correction factor, *CFI* comparative fit index, *RMSEA* root mean square of approximation, *SRMR* standardized root mean square residual

Satorra-Bentler χ^2, all $p < .001$

mean and intercept estimates based on the resulting model are also given in Supplementary Information (Tables SI.4, SI.5).

The number of non-equivalent intercepts ranged from 6 to 12 per item. Some of the intercepts revealed meaningful patterns. For instance, non-equivalence of the intercept of item three (achieving in life) was more often found in Asian, collectivistic cultures. Non-

Table 5 Comparison of observed country mean scores and latent factor estimates

	Observed score			MGCFA estimates		
	M	SD	Rank	M	Variance	Rank
Romania	8.07	1.33	1	.07	.92	1
India	7.67	1.36	2	−.27	1.23	2
Chile	7.25	1.37	8	−.38	1.04	3
Slovakia	7.43	1.16	5	−.40	.75	4
Panama	7.31	1.69	6	−.46	2.05	5
Spain	7.15	1.20	10	−.47	.64	6
Hungary	7.58	1.29	3	−.50	1.10	7
Puerto Rico	7.29	1.62	7	−.54	1.55	8
Armenia	7.52	1.55	4	−.55	1.44	9
Brazil	7.13	1.46	11	−.59	1.25	10
Israel	7.15	1.31	9	−.67	1.00	11
Portugal	6.89	1.27	16	−.87	.93	12
Bulgaria	7.02	1.54	13	−.92	1.38	13
South Africa	6.99	1.40	14	−1.01	1.09	14
Vietnam	6.78	1.42	19	−1.12	1.17	15
Estonia	7.10	1.21	12	−1.13	.82	16
Poland	6.91	1.54	15	−1.14	1.27	17
Serbia	6.80	1.65	18	−1.14	1.63	18
Indonesia	6.45	1.60	20	−1.25	1.56	19
UK	6.85	1.38	17	−1.29	1.20	20
Russia	6.40	1.64	22	−1.30	1.74	21
South Korea	6.15	1.34	23	−1.61	1.07	22
Nepal	6.44	1.41	21	−1.62	1.08	23
Japan	5.38	1.45	26	−2.06	1.41	24
Kenya	5.86	2.15	24	−2.26	3.03	25
Iran	5.72	1.61	25	−2.33	1.28	26

equivalence of the intercept of item six (feeling part of your community) was typically found in Latin American countries, but not in post-Communist ones.

The estimates of latent factor means and variances obtained from the model are presented in Table 5. The latent factor means were highly correlated with the observed means ($r = .95$).

To cross-validate the model, we performed the alignment procedure. The results are presented in Supplementary Information (Table SI.2). The alignment procedure has identified a smaller number of non-equivalent parameters, 1 loading and 37 intercepts, suggesting that the conventional approach is more conservative. Thirty-three of these intercepts were also identified as non-equivalent using the conventional approach. The latent factor means estimated using the alignment procedure were highly correlated with the observed means ($r = .98$) and with those obtained using the conventional approach ($r = .96$). These data indicate a fairly good convergence of the findings from the two procedures.

3.3 Invariance Analysis Across Gender

As our national samples differed in the gender distribution, we considered the possibility that some of the non-invariant parameters we found in the CFA analyses could be explained by gender. To gauge the contribution of gender to the measurement non-invariance, we performed a multi-group CFA for females and males in the combined sample, using country as a cluster variable to account for non-independence of observations within each country. The resulting fit indices and model comparison results are shown in Table 6. All three models showed good fit to the data. The difference between the nested models in terms of practical fit indices was below the thresholds suggested by Chen (2007), indicating that gender does not have any uniform effects on measurement invariance across countries.

4 Discussion

Our objective was to establish measurement invariance of the PWI-8 across 26 countries. We found that the PWI was unidimensional, with Cronbach's alphas indicating acceptable internal consistency in all countries. MGCFA confirmed that the basic construct structure of the PWI-8 is similar across groups. Although additional covariances between items can improve the fit in some countries, in most countries the fit of the theoretical model was very close to the acceptable range. The same was true for configural and metric-invariant MGCFA models. Although the difference between these two nested models was significant in terms of Chi square difference, the difference in practical fit indices was small, suggesting that the comparison of effects (e.g., correlations) obtained using the PWI in different languages across countries should not be biased by non-equivalent loadings. However, the poor fit of the scalar invariance model indicates that the comparison of raw scores between countries is impossible. On the other hand, the partial scalar invariance model indicated acceptable fit, allowing for a meaningful comparison of latent means across groups. Taken together, these findings suggest that it is possible to examine the predictors and correlates of the subjective well-being phenomenon using the PWI-8 across countries. The disagreement between the practical fit indices (ΔCFI vs. ΔRMSEA and

Table 6 Fit indices for multi-group confirmatory factor analysis (MGCFA) models of the PWI8 across gender

Model	S-B χ^2	df	CFI	RMSEA (90 % CI)	SRMR
Configural	218.12*	40	.968	.040 (.035–.046)	.028
Metric	231.90*	47	.967	.038 (.033–.043)	.031
Scalar	265.05*	54	.962	.038 (.033–.043)	.035
Difference tests	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
Metric versus configural	6.03	7	.001	.002	.003
Scalar versus metric	32.84*	7	.005	<.001	.004

CFI comparative fit index, *RMSEA* root mean square of approximation, *SRMR* standardized root mean square residual

Satorra-Bentler χ^2, * $p < .001$

ΔSRMR) can be explained by the fact that these fit indices are associated with the number of parameters in the model in different ways, and existing cut-off criteria (Chen 2007; Cheung and Rensvold 2002) are all based on simulations for two groups, where the number of parameters is much smaller. We could not find any simulation studies investigating optimal cut-off points for practical fit indices with a large number of groups.

Therefore, simulation studies investigating the effects of the number of groups on fit indices are necessary.

The results of analysis using an alignment approach based on the theoretical model suggest that despite the presence of some weak non-invariance of loadings and pronounced non-invariance of item intercepts in some countries, the observed mean scores are very similar to unbiased latent factor scores. Although removal of bias can slightly change the rank ordering of countries, these effects were only pronounced for a few countries (Kenya, Japan, Israel, and South Africa).

We also developed modified measurement models by introducing theoretically interpretable and strongly significant ($p < .001$) modification indices for some countries. However, in most cases these modification indices accounted for translation artefacts, which can be removed by improving certain translations of the instrument into other languages. Interestingly, in all countries where the English version was used instead of local languages (i.e., India, Iran, Kenya, Nepal, and South Africa), there was no need to introduce modifications and the unidimensional theoretical model was well fitted to the data.

The attempts to develop partial scalar invariance models using the older, manual approach and alignment approach resulted in different, although largely overlapping, sets of non-invariant parameters. Items referring to more objective realities, such as one's living standard and health, turned out to be more invariant, compared to the items referring to more subjective phenomena, such as one's spirituality or future security.

Finally, gender does not seem to contribute to non-invariance of loadings and intercepts in any uniform manner across countries. However, because our samples did not allow for the evaluation of gender invariance in each country separately, this analysis does not rule out the possibility of country-specific non-invariance associated with gender.

4.1 Limitations and Recommendation for Future Studies

The current report has several limitations: the use of student samples, the lack of several important cultures and countries (such as Chinese or American), and the overrepresentation of European countries. For practical reasons we used both online and paper–pencil surveys, and in some countries, the questionnaire was distributed in English, rather than in a native language, which may have led to increased measurement error. Also, the student samples were not representative of their respective countries, which precludes us from interpreting the substantial differences in the mean score estimates. Finally, although we decided in favour of using a combined item for measuring satisfaction with religion and spirituality, these constructs are not interchangeable (Piedmont and Friedman 2012). As this solution happened to work well both in CFA and MGCFA it could be used in cross-cultural comparisons; however, for further exploration of the importance of religion and spirituality as separate factors in shaping overall life satisfaction, two separate items should be used (see Casas et al. 2009; Sarriera et al. 2014).

In terms of specific recommendations, PWI researchers could use this tool in all of the countries included in the current study as an indicator of general life satisfaction, measured by a combination of domain-specific items. The small number of non-invariant loadings

that we found suggests that satisfaction in these domains contributes more or less equally to overall life satisfaction. However, we have found some consistently repeating cultural differences, which could be further explored. For instance, in the South Korean sample we have found evidence in favour of an additional factor representing concern about financial success, suggesting that in this population life satisfaction might be somewhat affected by materialism. Differences in intercepts for some PWI items suggest that these items may have specific meaning in certain cultural contexts. For instance, lower intercept for item three, reflecting satisfaction with achievement in life, was typically found in collectivistic countries, indicating that individuals in such countries are somewhat less likely to admit satisfaction with their individual achievements. As collectivistic countries are typically "face-saving" cultures (Bond 1991), life achievement could mean that individuals just fit into their social environment, contrary to individualistic countries, where life achievements would mean developing unique characteristics. As PWI statements are typically very general (as they are aimed to represent broad life domains; see International Wellbeing Group 2013), the cultural meaning of these broad statements could be affected by cultural context. Our study provides some suggestions of where this may be the case (e.g., life achievement, feeling part of community).

5 Conclusion

The current report provided information about the possibility of cross-cultural research among university students based on PWI-8 scores, providing evidence of partial metric invariance allowing the cross-country comparison of effects, but not of group or individual raw scores. We also compared the results of different approaches to establishing unbiased factor means across countries. This provides valuable information on the further development of subjective well-being research in different cultural contexts. As the main goal of the International wellbeing group is to explore the importance of satisfaction with particular domains in shaping overall life satisfaction, our findings indicate that this research goal could be realised successfully in cross-cultural research.

References

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 1–14. doi:10.1080/10705511.2014.919210.

Baltatescu, S. (2014). Romanian-Hungarian cross-border region, Personal well-being index. In Alex C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 5598–5605). Netherlands: Springer.

Bond, M. H. (1991). *Beyond the Chinese face: Insights from psychology*. Oxford: Oxford University Press.

Bricker-Katz, G., Lincoln, M., & McCabe, P. (2009). A Life-time of stuttering: How emotional reactions to stuttering impact activities and participation in older people. *Disability and Rehabilitation, 31*, 1742–1752. doi:10.1080/09638280902738672.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: The Guilford Press.

Byrne, B. N. (2012). *Structural Equation modelling with mplus*. New York: Routledge.

Casas, F., González, M., Figuer, C., & Malo, S. (2009). Satisfaction with spirituality and with religion, and personal well-being among Spanish adolescents and young university students. *Applied Research in Quality of Life, 4*(1), 23–45.

Casas, F., Bello, A., González, M., & Aligué, M. (2012). Personal well-being among Spanish adolescents. *Journal of Social Research and Policy, 3*, 19–45. doi:10.1007/s11482-009-9066-x.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504.

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

Cummins, R. A., Eckersley, R., Pallant, J., Van Vugt, J., & Misajon, R. (2003). Developing a national index of subjective wellbeing: The Australian unity wellbeing index. *Social Indicators Research, 64*, 159–190. doi:10.1023/A:1024704320683.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology, 40*, 55–75. doi:10.1146/annurev-soc-071913-043137.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542–575.

Diener, E., Horowitz, J., & Emmons, R. A. (1985). Happiness of the very wealthy. *Social Indicators Research, 16*, 263–274.

Engel, L., & Cummins, R. A. (2011). Impact of dose adjustment for normal eating in Australia (OzDAFNE) on subjective wellbeing, coping resources and negative affect in adults with type 1 diabetes: A prospective comparison study. *Diabetes Research and Clinical Practice, 91*, 271–279. doi:10.1016/j.diabres.2010.11.023.

Forjaz, F. M., Prieto-Flores, M. E., Ayala, A., Rodriguez-Blazquez, C., Fernandez-Mayoralas, G., Rojo-Perez, F., et al. (2011). Measurement properties of the community wellbeing index in older adults. *Quality of Life Research, 20*, 733–743. doi:10.1007/s11136-010-9794-2.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling, 6*, 1–55.

Huntington, S. (1996). *The clash of civilizations and remaking of world order*. New York: Touchstone.

International Wellbeing Group (2006). *Personal wellbeing index*. Melbourne: Australian Centre on Quality of Life, Deakin University. Retrieved from http://www.deakin.edu.au/research/acqol/instruments/wellbeing_index.htm.

International Wellbeing Group (2013). *Personal wellbeing index: 5th edition*. Melbourne: Australian Centre on Quality of Life, Deakin University. Retrieved from http://www.deakin.edu.au/research/acqol/instruments/wellbeingindex/index.php.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cut-off criteria: What did they really say? *Organizational Research Methods, 9*, 202–220. doi:10.1177/1094428105284919.

Lau, A. L. D., Cummins, R. A., & McPherson, W. (2005). An Investigation into the cross-cultural equivalence of the personal wellbeing index. *Social Indicators Research, 72*, 403–430.

Norris, P., & Inglehart, R. (2004). *Sacred and secular: Religion and politics worldwide*. New York: Cambridge University Press.

Oishi, S. (2010). Culture and well-being: Conceptual and methodological issues. In E. Diener, J. F. Helliwell, & D. Kahneman (Eds.), *International differences in well-being* (pp. 34–69). Oxford: Oxford University Press.

Piedmont, R. L., & Friedman, P. (2012). Spirituality, religiosity and subjective well-being. In K. C. Land, A. C. Michalos, & M. J. Sirgy (Eds.), *Handbook of social indicators and quality of life research* (pp. 313–329). Netherlands: Springer. doi:10.1007/978-94-007-2421-1.

Ponizovsky, Y., Dimitrova, R., Schachner, M. K., & Van de Schoot, R. (2013). The satisfaction with life scale: Measurement invariance across immigrant groups. *European Journal of Developmental Psychology, 10*, 526–532. doi:10.1080/17405629.2012.707778.

Sarriera, J. C., Casas, F., Alfaro, J., Bedin, L., Strelhow, M. R., Abs, D., et al. (2014). Psychometric properties of the personal wellbeing index in Brazilian and Chilean adolescents including spirituality and religion. *Psychology/Psicologia Reflexão e Crítica, 27*, 710–719. doi:10.1590/1678-7153.201427411.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Sirgy, M. J. (2012). *The psychology of quality of life. Hedonic well-being, life satisfaction, and eudaimonia*. New York: Springer. doi:10.1007/978-94-007-4405-9.

Tomyn, A. J., Norrish, J. M., & Cummins, R. A. (2011). The subjective wellbeing of indigenous Australian adolescents: validating the personal wellbeing index—school children. *Social Indicators Research, 103*, 405–411. doi:10.1007/s11205-011-9970-y.

Tucker, K. L., Ozer, D. J., Lyubomirsky, S., & Boehm, J. K. (2006). Testing for measurement invariance in the satisfaction with life scale: A comparison of Russians and North Americans. *Social Indicators Research, 78*, 341–360.

Werner, S. (2012). Subjective well-being, hope, and needs of individuals with serious mentalillness. *Psychiatry Research, 196*, 214–219. doi:10.1016/j.psychres.2011.10.012.

Whisman, M. A., & Judd, C. M. (2016). A cross-national analysis of measurement invariance of the satisfaction with life scale. *Psychological Assessment, 28*(2), 239–244.