



Evaluating the Construct Validity of Ryff's Scales of Psychological Well-Being Using Exploratory Structural Equation Modeling

Journal of Psychoeducational Assessment

2017, Vol. 35(6) 633–638

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0734282916652756

journals.sagepub.com/home/jpa



Hsien-Yuan Hsu¹, Tze-Li Hsu², KoFan Lee¹, and Lori Wolff¹

Abstract

The purpose of this study was to evaluate the construct validity of Ryff's Scales of Psychological Well-Being (SPWB) using exploratory structural equation modeling (ESEM). The data were drawn from the national survey of Midlife in the United States conducted during 1994 and 1995. Measurement models assuming different number of factors (1–6 factors) and considering the effect of negatively wording items were specified and compared to determine optimal number of underlying factors. The discriminant validity was assessed following Farrell's suggestions. The results showed the discriminant validity was questionable due to five indicators with considerable cross-loadings.

Keywords

construct validity, exploratory structural equation modeling, psychological well-being

Ryff's Scales of Psychological Well-Being (SPWB) embrace six dimensions of positive functioning (Ryff, 1989)—autonomy (*AU*), environmental mastery (*EM*), personal growth (*PG*), positive relations (*PR*), purpose in life (*PL*), and self-acceptance (*SA*). The stability of these dimensions was supported by substantive studies across the last two decades (Ryff, 2014). Previous researchers have relied on exploratory factor analysis (EFA; Kafka & Kozma, 2002) or confirmatory factor analysis (CFA; Lindfors, Berntsson, & Lundberg, 2006; Ryff & Keyes, 1995; Springer & Hauser, 2006) to investigate SPWB's factorial structure. However, EFA and CFA are sometimes inappropriate to multidimensional scales; therefore, the findings of aforementioned studies can be misleading (Marsh, Liem, Martin, Morin, & Nagengast, 2011). More specifically, EFA assumes unique variances are independent, while unique variances of negatively wording items of SPWB are dependent (Springer & Hauser, 2006). CFA does not assume independent unique variances, but researchers often constrain cross-loadings to zero, resulting in poor goodness-of-fit to the data and inflated inter-factor correlations that can undermine discriminant validity (Asparouhov & Muthén, 2009; Hsu, Skidmore, Li, & Thompson, 2014; Marsh et al., 2011).

¹University of Mississippi, University, MS, USA

²Sam Houston State University, Huntsville, TX, USA

Corresponding Author:

Hsien-Yuan Hsu, Department of Leadership and Counselor Education, University of Mississippi, 49 Rebel Drive, 103 Guyton Hall, University, MS 38677, USA.

Email: hhsu2@olemiss.edu

Exploratory structural equation modeling (ESEM) is a more appropriate method for evaluating the construct validity of SPWB. ESEM integrates positive features of EFA and CFA (Asparouhov & Muthén, 2009; Marsh, Morin, Parker, & Kaur, 2014). ESEM tends to show a better model–data fit and more accurate estimates of inter-factor correlations compared with CFA (Marsh et al., 2011; Marsh et al., 2014). ESEM has been recently applied to investigate the construct validity of multidimensional scales in psychological (e.g., Booth & Hughes, 2014; Vazsonyi, Ksinan, Mikuska, & Jiskrova, 2015) and educational (e.g., Caro, Sandoval-Hernández, & Lüdtke, 2014; Marsh et al., 2013) studies. The purpose of this study was to evaluate the construct validity of SPWB using ESEM. Specifically, models with different number of factors (1-6 factors) and a method factor associated with negatively wording items were specified and compared to determine an optimal number of factors. We then assessed the discriminant validity following Farrell's (2010) suggestions.

Method

Data

The data were drawn from the national survey of Midlife in the United States (MIDUS) conducted during 1994 and 1995 ($N = 3,487$). Participants who did not complete SPWB were excluded, resulting in 3,014 participants, of which 48.5% were males, 87.72% Caucasian, and the average age was 47.07 years ($SD = 13.19$ years).

Measures

In MIDUS, each dimension is measured by three 7-point Likert-type items (see the appendix), where at least one item is negatively worded. Following Springer and Hauser's (2006) study, the category *do not know* was recoded as missing, and the remaining categories were recoded from 1 to 6, where a larger number represented a stronger agreement with the items' description. Responses on the negatively wording items were reverse coded. Our preliminary analysis showed that those analytical items were mildly skewed (skewness ranged from -0.032 to -1.855), which can be appropriately handled by the robust weighted least squares estimator described as following (Lei, 2009).

Data Analysis

Seven models were specified—*Model1* to *Model6* assumed one to six factors, respectively, while *Model7* added a method factor associated with negatively worded items to *Model6*. Data were analyzed in ordinal-scale using a robust weighted least squares estimator in Mplus, namely, a mean- and variance-adjusted weighted least squares (WLSMV) estimator. We adopted the pairwise deletion approach accommodated in WLSMV estimator to handle missing data (Asparouhov & Muthén, 2010). The rotation method of ESEM was Geomin (Asparouhov & Muthén, 2009). The model fit was evaluated by χ^2 statistics and fit indices—comparative fit index (CFI), Tucker–Lewis index (TLI), and root mean square error of approximation (RMSEA) with cutoff values (e.g., CFI, TLI ≥ 0.95 ; RMSEA ≤ 0.06) recommended by Hu and Bentler (1999). χ^2 difference ($\Delta\chi^2$) tests were conducted to compare relative fit of nested models that are differed in number of factors (e.g., five-factor model vs. six-factor model). The 90% confidence interval (CI) of RMSEA was also examined for model comparison (Preacher, Zhang, Kim, & Mels, 2013).

Discriminant validity. We compared the average variance extracted (AVE) of each factor with the shared variance between factors (Farrell, 2010). The factor's AVE is the average amount of squared factor loadings of its indicators, while shared variance is the squared correlation between

Table 1. Correlations (SE) and Shared Variances Between Factors of Model7.

Factor	1	2	3	4	5	6
1. Autonomy	—	.238	.135	.032	.020	.127
2. Environmental mastery	.488 (0.020)*	—	.184	.081	.044	.367
3. Personal growth	.367 (0.023)*	.429 (0.023)*	—	.024	.055	.114
4. Positive oing relations	.180 (0.034)*	.284 (0.037)*	.156 (0.038)*	—	.009	.174
5. Purpose in life	.140 (0.030)*	.210 (0.033)*	.235 (0.040)*	-.093 (0.098)	—	.208
6. Self-acceptance	.356 (0.022)*	.606 (0.025)*	.337 (0.024)*	.417 (0.067)*	.456 (0.050)*	—

Note. Lower diagonal contains correlations (SE) between factors; upper diagonal contains shared variances.

* $p < .05$.

any two factors. Discriminant validity is supported when the AVE for any two factors is larger than the maximum shared variance.

Results and Discussion

Model1 had the worst goodness of fit, $\chi^2(df) = 8,203.860 (135)$, $p < .05$; CFI = 0.718, TLI = 0.680, RMSEA = 0.141. Models with an increasing number of factors (*Model2-Model6*) showed better goodness of fit. *Model6* performed statistically better than all other five models, $\chi^2(df) = 680.824 (60)$, $p < .05$, CFI = 0.978, TLI = 0.945, RMSEA = 0.059. These results indicated that there were six substantive factors. Compared with *Model6*, *Model7* added an additional method factor to account for negatively wording items. This model demonstrated an adequate fit, $\chi^2(df) = 185.836 (52)$, $p < .05$, CFI = 0.994, TLI = 0.981, RMSEA = 0.034, and fit significantly better than *Model6* with $\Delta\chi^2(df) = 357.961 (8)$, $p < .05$ and non-overlapping 90% CI of RMSEA, *Model6* = [0.055, 0.063]; *Model7* = [0.030, 0.039]. Therefore, *Model7* was chosen as the optimal model. Table 1 presents the correlations (lower diagonal) and shared variances (upper diagonal) between factors in *Model7*.

Our results supported that there were six substantive factors, which is in line with previous studies using CFA (e.g., Lindfors et al., 2006; Ryff & Keyes, 1995; Springer & Hauser, 2006) but against one study using EFA (Kafka & Kozma, 2002). However, our study revealed smaller magnitudes of inter-factor correlation compared with previous CFA studies. For instance, in a comparable study analyzing the same data as the current study, Springer and Hauser (2006) found that all inter-factor correlations were above .763 (see Table 7 in their study) after controlling for method effects (e.g., negative wording items). Other studies using CFA also reported similar findings—high inter-factor correlations of SPWB (e.g., Lindfors et al., 2006). The above findings are echoed by prior studies indicating inappropriately forcing cross-loadings to zero led to inflated inter-factor correlations (Asparouhov & Muthén, 2009; Marsh et al., 2009). Note that the factor structure of SPWB derived from the ESEM in the current study cannot be overgeneralized to other populations, and future studies are needed to cross-validate our findings.

Discriminant Validity

The estimated inter-factor correlations from *Model7* were all positive except for correlation between *PR* and *PL* ($-.093$, $p > .05$). The magnitudes of factor correlation ranged from $-.093$ to $.606$, resulting in shared variances ranging from $.009$ to $.367$. Table 2 presents pattern coefficients of each indicator and AVE (ranging from $.184$ [*PR*] to $.536$ [*SA*]) for each factor. Three factors

Table 2. Factor Loadings (SE) of Model7.

Items	Factor						AVE
	AU	EM	PG	PR	PL	SA	
AU1 ^a	.356 (0.024)	-.020 (0.030)	-.128 (0.028)	-.036 (0.031)	.218 (0.035)	.047 (0.036)	.467
AU2	.864 (0.028)	-.008 (0.016)	.027 (0.015)	-.014 (0.014)	.028 (0.010)	.018 (0.018)	
AU3	.727 (0.027)	.042 (0.024)	.000 (0.013)	.036 (0.020)	-.065 (0.023)	-.023 (0.023)	
EM1	.023 (0.017)	.637 (0.059)	.023 (0.014)	-.048 (0.017)	-.037 (0.020)	.220 (0.056)	.365
EM2 ^a	.035 (0.020)	.162 (0.048)	-.016 (0.014)	-.404 (0.072)	.042 (0.045)	.402 (0.080)	
EM3	.017 (0.017)	.815 (0.034)	.032 (0.018)	.046 (0.019)	.024 (0.015)	-.014 (0.011)	
PG1	-.025 (0.013)	-.019 (0.016)	.932 (0.028)	-.015 (0.016)	-.002 (0.016)	.001 (0.016)	.448
PG2	.039 (0.017)	.038 (0.022)	.656 (0.028)	.008 (0.017)	.065 (0.029)	.174 (0.034)	
PG3 ^a	-.027 (0.017)	.012 (0.020)	.209 (0.045)	-.015 (0.019)	.695 (0.030)	.025 (0.023)	
PR1 ^a	.024 (0.015)	.020 (0.016)	-.107 (0.027)	.526 (0.062)	.464 (0.136)	.060 (0.042)	.184
PR2	.097 (0.026)	.067 (0.033)	.391 (0.030)	.296 (0.031)	-.022 (0.035)	-.002 (0.023)	
PR3 ^a	-.011 (0.016)	.036 (0.024)	.048 (0.020)	.434 (0.050)	.442 (0.092)	.038 (0.030)	
PL1 ^a	-.057 (0.025)	-.008 (0.029)	-.028 (0.027)	.093 (0.029)	.539 (0.031)	-.016 (0.030)	.221
PL2	.006 (0.020)	.202 (0.036)	.018 (0.019)	.106 (0.026)	.063 (0.030)	.448 (0.037)	
PL3 ^a	.071 (0.026)	-.185 (0.042)	.102 (0.042)	-.045 (0.023)	.607 (0.030)	-.048 (0.032)	
SA1	-.004 (0.011)	-.066 (0.021)	.023 (0.011)	.033 (0.015)	-.191 (0.068)	1.010 (0.051)	.536
SA2	.215 (0.023)	.058 (0.032)	.115 (0.023)	.114 (0.028)	.010 (0.016)	.420 (0.032)	
SA3 ^a	-.019 (0.016)	.048 (0.036)	-.035 (0.017)	-.214 (0.059)	.113 (0.048)	.642 (0.067)	

Note. Five indicators with unexpected low main-loadings but high cross-loadings are highlighted by a shaded box. AU = autonomy; EM = environmental mastery; PG = personal growth; PR = positive relations with others; PL = purpose in life; SA = self-acceptance; AVE = average variance extracted.

^aNegatively worded items.

(*EM*, *PR*, and *PL*) had AVE lower than the maximum shared variance, which can be an evidence of lack of discriminant validity of SPWB (Fornell & Larcker, 1981).

Lack of discriminant validity may be attributable to five problematic indicators (*EM2*, *PG3*, *PR2*, *PR3*, and *PL2*, highlighted in Table 2). For example, *PG3* had a small loading on the target factor *PG* (.209) but a large cross-loading on factor *PL* (.695). The variances of these problematic indicators were not mainly explained by the corresponding theory-driven factors. Similarly, Clarke, Marshall, Ryff, and Wheaton (2001) also identified four problematic SPWB indicators, and three of them (i.e., *EM2*, *PR2*, and *PL2*) were also found in the current study. As suggested by Clarke et al., all problematic indicators need to be modified or replaced.

Taken together, the factorial structure of SPWB determined by ESEM included six underlying factors with small to moderate inter-factor correlations, a method factor associated with negatively worded item, and five problematic indicators. Problematic indicators led to lower AVE of three

Appendix

Wording and Order of Items in MIDUS.

Construct	Item	Description
AU	AU1	"I tend to be influenced by people with strong opinions." (N)
	AU2	"I have confidence in my own opinions, even if they are different from the way most other people think."
	AU3	"I judge myself by what I think is important, not by the values of what others think is important."

(continued)

Appendix (continued)

Construct	Item	Description
EM	EM1	"In general, I feel I am in charge of the situation in which I live."
	EM2	"The demands of everyday life often get me down." (N)
	EM3	"I am quite good at managing the many responsibilities of my daily life."
PG	PG1	"I think it is important to have new experiences that challenge how you think about yourself and the world."
	PG2	"For me, life has been a continuous process of learning, changing, and growth."
	PG3	"I gave up trying to make big improvements or changes in my life a long time ago." (N)
PR	PR1	"Maintaining close relationships has been difficult and frustrating for me." (N)
	PR2	"People would describe me as a giving person, willing to share my time with others."
	PR3	"I have not experienced many warm and trusting relationships with others." (N)
PL	PL1	"I live life one day at a time and do not really think about the future." (N)
	PL2	"Some people wander aimlessly through life, but I am not one of them."
	PL3	"I sometimes feel as if I have done all there is to do in life." (N)
SA	SA1	"When I look at the story of my life, I am pleased with how things have turned out."
	SA2	"I like most parts of my personality."
	SA3	"In many ways I feel disappointed about my achievements in life." (N)

Note. MIDUS = Midlife in the United States; AU = autonomy; EM = environmental mastery; PG = personal growth; PR = positive relations with others; PL = purpose in life; SA = self-acceptance; N = negatively wording items.

factors (*EM*, *PR*, and *PL*) and contributed to the lack of discriminant validity of SPWB. Researchers should be cautious in interpreting the scores measured by these problematic indicators.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Asparouhov, T., & Muthén, B. O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397-438.
- Asparouhov, T., & Muthén, B. O. (2010). *Weighted least squares estimation with missing data*. Los Angeles, CA: Muthén & Muthén.
- Booth, T., & Hughes, D. J. (2014). Exploratory structural equation modeling of personality data. *Assessment, 21*, 260-271. doi:10.1177/1073191114528029
- Caro, D., Sandoval-Hernández, A., & Lüdtke, O. (2014). Cultural, social, and economic capital constructs in international assessments: An evaluation using exploratory structural equation modeling. *School Effectiveness and School Improvement, 25*, 433-450. doi:10.1080/09243453.2013.812568
- Clarke, P. J., Marshall, V. W., Ryff, C. D., & Wheaton, B. (2001). Measuring psychological well-being in the Canadian study of health and aging. *International Psychogeriatrics, 13*, 79-90. doi:10.1017/S1041610202008013
- Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research, 63*, 324-327. doi:10.1016/j.jbusres.2009.05.003

- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50.
- Hsu, H.-Y., Skidmore, S. T., Li, Y., & Thompson, B. (2014). Forced zero cross-loading misspecifications in measurement component of structural equation models: Beware of even "small" misspecifications. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 10, 138-152. doi:10.1027/1614-2241/a000084
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Kafka, G. J., & Kozma, A. (2002). The construct validity of Ryff's Scales of Psychological Well-Being (SPWB) and their relationship to measures of subjective well-being. *Social Indicators Research*, 57, 171-190.
- Lei, P.-W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality & Quantity*, 43, 495-507. doi:10.1007/s11135-007-9133-z
- Lindfors, P., Berntsson, L., & Lundberg, U. (2006). Factor structure of Ryff's psychological well-being scales in Swedish female and male white-collar workers. *Personality and Individual Differences*, 40, 1213-1222.
- Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment*, 29, 322-346. doi:10.1177/0734282911406657
- Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory Structural Equation Modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10, 85-110. doi:10.1146/annurev-clinpsy-032813-153700
- Marsh, H. W., Muthén, B., Asparouhov, A., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, 16, 439-476. doi:10.1080/10705510903008220
- Marsh, H. W., Vallerand, R. J., Lafrenière, M. K., Parker, P., Morin, A. J. S., Carbonneau, N., . . . Paquet, Y. (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, 796-809. doi:10.1037/a0032573
- Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, 48, 28-56. doi:10.1080/00273171.2012.710386
- Ryff, C. D. (1989). Beyond Ponce de Leon and life satisfaction: New directions in quest of successful ageing. *International Journal of Behavioral Development*, 12, 35-55.
- Ryff, C. D. (2014). Psychological well-being revisited: Advances in the science and practice of eudaimonia. *Psychotherapy and Psychosomatics*, 83, 10-28. doi:10.1159/000353263
- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69, 719-727.
- Springer, K. W., & Hauser, R. M. (2006). An assessment of the construct validity of Ryff's Scales of Psychological Well-Being: Method, mode, and measurement effects. *Social Science Research*, 35, 1080-1102.
- Vazsonyi, A. T., Ksinan, A., Mikuska, J., & Jiskrova, G. (2015). The Big Five and adolescent adjustment: An empirical test across six cultures. *Personality and Individual Differences*, 83, 234-244. doi:10.1016/j.paid.2015.03.049