KARI L. TUCKER, DANIEL J. OZER, SONJA LYUBOMIRSKY
and JULIA K. BOEHM

TESTING FOR MEASUREMENT INVARIANCE IN THE SATISFACTION WITH LIFE SCALE: A COMPARISON OF RUSSIANS AND NORTH AMERICANS

ABSTRACT. This study examined the comparability of Satisfaction With Life Scale (SWLS) [Diener, Emmons, Larsen, & Griffin, 1985, *Social Indicators Research*, 34: 7–32] scores across U.S. and Russian student and community groups. Criteria for weak measurement invariance were met when comparing U.S. and Russian groups (combining student and community samples). Criteria for weak and strong measurement invariance were met when comparing the U.S. and Russian student samples. However, when comparing the U.S. and Russian community samples, the results showed a significant statistic for a baseline model, indicating a lack of comparability across samples. The costs of failing to meet criteria for weak, strong, and strict measurement invariance are discussed.

KEY WORDS: culture, life satisfaction, measurement invariance, subjective well-being

INTRODUCTION

People differ dramatically in what they require for a satisfying life. To some, long walks on the beach, time with family, and high marks in school are the most important sources of overall life satisfaction, while others consider personal freedom, a comfortable income, and robust health to be most critical. Given that sources of life satisfaction vary widely among individuals, it would be natural for them to vary across cultures and subcultures. One might imagine, for example, that males and females, the young and old, and individuals living in prosperous countries and those living in developing countries would derive life satisfaction from different sources, depending on their life circumstances.

Research supports the notion that what may be important for some individuals or cultures may not be as important for others, showing, for example, that income (Veenhoven, 1991), self-esteem (Diener and Diener, 1995; Diener et al., 1995; Oishi et al., 1999), number of friends (Sam, 2001), and satisfaction with education, material wealth, home life, and health (Diener et al., 1995; Oishi et al., 1999; Sam, 2001) correlate differently with life satisfaction in different groups. Thus, using satisfaction with specific domains (e.g., income, education, and health) as indicators in assessing overall life satisfaction may pose problems for researchers wishing to compare groups.

Diener and his colleagues (Diener et al., 1985) resolved this issue by constructing a scale that measures life satisfaction more generally. That is, rather than assessing global life satisfaction by summing up respondents' satisfactions with a variety of specific sources (e.g., recreation or self-esteem), these authors assess life satisfaction subjectively by examining people's perceptions of life as a whole. With this subjective approach, respondents may use whatever sources they choose for evaluating how satisfied they are with their lives overall. Researchers are then able to compare groups using items that are presumably free from culturally specific definitions, as well as from individuals' varying criteria of life satisfaction (Pavot and Diener, 1993).

The Satisfaction With Life Scale (SWLS; Diener et al., 1985) was originally developed to circumvent problems inherent in previous scales, which were either composed of single items, narrowly focused on geriatric populations, or did not strictly measure the judgmental aspects of life satisfaction (Diener et al., 1985). Since its inception, the SWLS has been found to represent a single factor (e.g., Diener et al., 1985; Arrindell et al., 1991, 1999; Pavot et al., 1991; Shevlin and Bunting, 1994; Atienza et al., 2003) and to demonstrate adequate internal consistency (Cronbach's α's ranging from 0.79 to 0.89) and stability across time (r's of 0.84 for 1 month and 0.54 for 4 years; Pavot and Diener, 1993) and occasions (Eid and Diener, 2004).

The positive psychometric properties of the SWLS, combined with its subjective approach, have invited numerous direct group comparisons. For example, researchers have compared the SWLS scores of many diverse groups, including British male and female students (Shevlin et al., 1998), Dutch adults of varying marital status

(Arrindell et al., 1991), adults from individualist and collectivist nations (Sam, 2001; Schimmack et al., 2002), and young and old individuals from different cultural backgrounds – e.g., French-Canadian (Blais et al., 1989; Pavot and Diener, 1993), Russian (Balatsky and Diener, 1993), Chinese, and Korean (Pavot and Diener, 1993; Diener et al., 1995). When scale means from such different groups are reported, it is generally presumed that the scores are directly comparable. However, the comparability of the SWLS across diverse groups has been more an assumed fact than a demonstrated outcome (for an exception, see Shevlin et al., 1998). For example, divergences in interpretation can arise when items on a scale do not carry similar connotations across groups. Participants may vary in their interpretations of certain words, their understanding of the intended meaning of a question, or their conceptualization of an entire scale due to differences in language or in cultural assumptions (Veenhoven, 1996). Under these circumstances, the items of the scale do not similarly represent the same latent construct (e.g., life satisfaction) across groups, and, as a result, the accuracy of interpretations about group differences on the latent construct is compromised.

Problems in the interpretability of a scale score arise when groups have equal standing at the latent level, but have unequal expected observed scores (Dragow, 1987). In other words, the scale is not measuring similarly across groups. When a scale is not measuring similarly, measurement is biased – that is, measured group differences do not reflect real differences at the latent level. In order to make group comparisons, researchers need to establish measurement invariance. Measurement invariance is achieved when parameters of the measurement model are equivalent across groups (Meredith, 1993; see also Widaman and Reise, 1997). When each measured item is a function of a single latent trait, the item parameters include a regression coefficient, an intercept, and error variance. Structural equation models may then be used to systematically test all three components to verify their equivalence across groups.

Establishing Measurement Invariance

A first level of measurement invariance is reached when there is equivalence of the regression coefficients across groups. Researchers

have recognized the need to check for equivalence across groups at this level by examining factor loadings (for examples in well-being research, see Carp and Carp, 1983; Liang and Bollen, 1985; Liang et al., 1986, 1987; Blais et al., 1989; Arrindell et al., 1991; Pavot et al., 1991; Shevlin and Bunting, 1994). In structural equation modeling, factor loadings are the regression coefficients for predicting item responses from the latent variable. Equivalence of factor loadings (i.e., regression coefficients) demonstrates that the latent variable is related to the items in the same way across groups (Reise et al., 1993). Establishing equivalence at this level is important because each item that is chosen to measure a latent variable should equally measure that latent variable across groups. Thus, structural equation modeling offers an objective test for establishing the equivalence of the factor loadings for each item across groups. Testing the equivalence of factor loadings across groups has been termed "weak" measurement invariance (Meredith, 1993) because it is only the first level of measurement invariance for the linear model.

A second, "strong" level of measurement invariance holds when, in addition to item slopes, there is equivalence of item intercepts across groups. Item intercepts are the values of the observed scores when the latent trait is zero. In structural equation modeling, the observed mean differences at the item level should reflect the mean difference at the latent level; otherwise, measurement bias is evident. Strong measurement invariance requires that mean differences at the observed level reflect the mean group difference at the latent level. In addition, weak measurement invariance is a necessary precondition for strong measurement invariance. To compare group means on manifest variables, strong measurement invariance must be established.

A third, "strict" level of measurement invariance is achieved when, in addition to item slopes and intercepts, there is equivalence of the item residuals across groups (Drasgow, 1984, 1987; Byrne et al., 1989). Item residuals are the unique item variances – that is, item variance not attributable to the latent variable. Demonstrating strict measurement invariance means that unique item variances are constant across groups. Because unique variances affect the magnitude of correlations among observed variables, strict measurement invariance must be established in order to compare correlations among manifest variables across groups.

Present Study

The purpose of the present study was to evaluate whether the SWLS measures similarly across two cultural groups. The United States and Russia were selected for comparison because of their strikingly different norms regarding the expression of positive affect (Lyubomirsky, 1997). Claimed as a right in the Declaration of Independence, personal happiness and life satisfaction play a central role in the daily social life and intellectual discourse of the United Sates. The majority of U.S. respondents rate life satisfaction as very important (Triandis et al., 1990; Diener et al., 1995) and report thinking about their personal happiness at least once every day (Freedman, 1978). By contrast, Russians are less likely to believe that the ideal life is worth pursuing, compared to their U.S. peers (Lyubomirsky, 1997). Russian social life and language are rich in resources for expressing *negative* affect (Wierzbicka, 1994), and Russians appear to be relatively more concerned with the sharing of misfortune. Indeed, the expression of life satisfaction and success is often perceived to risk inviting envy, resentment, suspicion, or the "evil eye" (Smith, 1990). A historical distrust of the system, combined with hopelessness, lack of control, and suspicions that anyone who is very satisfied with life must have used "crooked" means, steers Russians away from expressing positive feelings to others to avoid inviting negative social comparisons (Balatsky and Diener, 1993).

We expected cultural differences between U.S. and Russian individuals to be magnified in older populations as compared to younger populations (e.g., university students). University students share many more goals and experiences with one another – even across continents – than do working adults. In contrast, working community members, who belong to a different cohort, are older and more likely to embody the traditions and values of their culture. Indeed, given the tremendous recent social, economic, and political changes in their country, the values and attitudes of young people in Russia are likely to conflict with those of the older generation. Because the Russian student participants in this study spent their formative years (i.e., since age 11 or 12) in the post-Soviet regime, without Communism, they are much more oriented toward the West, confident, and hopeful about their lives compared to their parents and older peers. Indeed, Russian students report surprisingly similar attitudes as their U.S. counterparts

(cf. Grob et al., 1996; Lyubomirsky, 1997; Ryan et al., 1999). There-fore, because of possible differences in values and attitudes regarding life satisfaction due to age or life experiences, measurement invariance for the SWLS was tested across student and community-based samples from the United States and Russia. We expected to observe at least strong measurement invariance across the U.S. and Russian student samples, but not across the adult community-based samples.

METHODS

Participants

Responses were collected from a U.S. sample, comprised of 86 students attending a public university (59% female) and 62 commu-nity members (64% female) residing in Southern California, and a Russian sample, comprised of 66 students attending Moscow State University (73% female) and 63 community members (65% female) residing in Moscow, Russia, and its suburbs. The U.S. student sample ranged in ages from 17 to 24 ($M = 18.8$); the U.S. community sample ranged in ages from 18 to 62 ($M = 41.2$); the Russian stu-dent sample ranged in ages from 17 to 28 ($M = 19.8$); and the Russian community sample ranged in ages from 18 to 66 ($M = 38.8$). Both student samples were recruited in psychology classes and received either course credit or payment of $5.00 for their participation. The U.S. community sample was matched to the Russian community sample in sex, age, educational level, and occupational status.[1] U.S. community members either volunteered to participate in this study without compensation or were paid $10.00. The Russian community sample was recruited in their work places by native Russian speakers, and received $5.00 for their participation.

The Satisfaction With Life Scale

As part of a larger project, participants reported their levels of life satisfaction by completing the SWLS. The items on this scale are listed in Table I. Responses to the SWLS were made using 7-point Likert-type scales (1 = *strongly disagree*; 2 = *disagree*, 3 = *slightly dis-agree*; 4 = *neither agree nor disagree*; 5 = *slightly agree*; 6 = *agree*; 7 = *strongly agree*).

TABLE I

Means (Standard Deviations) For Each Item on the SWLS by Group

Item	Group					
	U.S (Comm)			Russian (Comm)		
	Students	Members	All	Students	Members	All
1. In most ways, my life is close to my ideal.	4.61 (1.38)	4.71 (1.52)	4.65 (1.44)	4.08 (1.53)	3.16 (1.29)	3.63 (1.49)
2. The conditions of my life are excellent.	4.81 (1.33)	4.79 (1.46)	4.80 (1.38)	4.20 (1.55)	3.08 (1.34)	3.65 (1.55)
3. I am satisfied with my life.	5.36 (1.08)	4.94 (1.64)	5.18 (1.36)	5.11 (1.43)	3.86 (1.29)	4.50 (1.50)
4. So far, I have gotten the important things I want in life.	5.17 (1.42)	4.94 (1.63)	5.07 (1.51)	4.26 (1.60)	3.30 (1.24)	3.79 (1.51)
5. If I could live my life over, I would change almost nothing.	4.64 (1.67)	3.94 (1.75)	4.35 (1.74)	4.82 (1.71)	3.30 (1.53)	4.08 (1.79)

Note: SWLS = Satisfaction With Life Scale (Diener et al., 1985). Comm = Community.

Russian language equivalents of the English language statements were created by a back-translation method (Brislin, 1970), which required a speaker fluent in both English and Russian to translate the English version of the questionnaire into a Russian version (see Appendix). Then, a second bilingual individual translated the Russian version of the questionnaire back into English.

Specification and Interpretation of Measurement Invariance Models

Several steps were taken to test for measurement invariance. First, item covariances and means for each group were computed. Second, parameters in the baseline, weak, strong, and strict models were estimated using the guidelines and procedures outlined by Widaman and Reise (1997). Specifically, the baseline model tested whether the five items on the scale loaded on one factor for both groups, and represented a sufficient model of the data (Byrne et al., 1989). In this model, parameters – factor loadings, factor variance, and unique item variances – were freely estimated for each group with minimal constraints. The weak measurement invariance model was specified the same as the baseline model except all five-factor loadings were constrained to equality across groups. The strong measurement invariance model was specified the same as the weak measurement invariance model except that the item intercepts were constrained to equality across groups. The strict measurement invariance model was specified the same as the model testing for strong measurement invariance except residual item variances were also constrained to equality across groups. Each model served as a basis for comparison with the preceding model (Byrne et al., 1989; Reise et al., 1993) because the baseline, weak, strong, and strict models are ''nested.''

Three fit indices were used in the present study: (1) the Tucker–Lewis Index (TLI; Tucker and Lewis, 1973), also called the non-normed fit index (NNFI; Bentler and Bonett, 1980); (2) the Goodness of Fit Index (GFI; Tanaka and Huba, 1985); and (3) the Root Mean Square Error of Approximation (RMSEA; Steiger and Lind, 1980). The TLI is a relative fit index that indicates the proportion of covariation among indicators explained by the specified model relative to a null model of independence. Values range from 0 (poor fit) to 1.0 (good fit), with values of 0.90 and above considered satisfactory (Reise et al., 1993). The GFI represents the amount of variance and covariance accounted

for by the model (Jöreskog and Sörbom, 1989), with values above 0.95 demonstrating good fit. Finally, the RMSEA is an absolute measure of the model's badness of fit (per degree of freedom), with a value of zero indicating that the model perfectly fits a set of data. RMSEA values of 0.05 indicate close fit, and values of 0.08 indicate reasonable fit (Browne, 1990).

RESULTS

SWLS item means and standard deviations for students and community members within the U.S. and Russian samples, as well as for the entire U.S. and Russian samples (students and community members combined) are shown in Table I. In addition, item covariances for these samples are shown in Table II.

Testing Measurement Invariance Across Groups

Comparing U.S. and Russian participants

Measurement invariance for the SWLS was first tested across the entire U.S. and Russian samples. As shown in Table III, the baseline

TABLE II

Item Covariances on the SWLS For Students (Community Members) Within the U.S. and Russian Samples

	Item 1	Item 2	Item 3	Item 4	Item 5
U.S. Sample					
Item 1	1.913 (2.307)				
Item 2	0.984 (1.679)	1.777 (2.137)			
Item 3	0.733 (1.898)	0.821 (2.172)	1.175 (2.683)		
Item 4	0.881 (1.160)	0.880 (1.092)	0.724 (1.307)	2.005 (2.650)	
Item 5	0.986 (1.375)	1.004 (1.333)	0.956 (1.551)	1.087 (1.750)	2.799 (3.077)
Russian Sample					
Item 1	2.347 (1.651)				
Item 2	0.691 (0.601)	2.406 (1.785)			
Item 3	1.437 (1.038)	0.763 (0.656)	2.033 (1.672)		
Item 4	1.349 (0.290)	0.688 (0.734)	1.096 (0.608)	2.563 (1.538)	
Item 5	1.413 (0.612)	0.466 (0.638)	0.926 (0.657)	0.787 (0.714)	2.952 (3.302)

Note: SWLS = The Satisfaction With Life Scale (Diener et al., 1985).

model demonstrated a non-significant chi-square statistic. This model served as a basis of comparison for the weak measurement invariance model. The results showed that criteria for weak measurement invariance were met. There was a non-significant decrease in model fit and good practical fit values. Thus, the items on the SWLS related to the latent variable in the same way across the two cultural groups. However, the results also showed that criteria for strong measure-

TABLE III

Parameter Estimates and Fit Indices for Testing the Baseline, Weak, and Strong Measurement Invariance Models in the U.S. and Russian Samples (Students and Community Members Combined)

Parameter	Baseline		Weak		Strong	
	U.S.	Russia	U.S.	Russia	U.S.	Russia
λ_{11}	1.06	1.06[b]	1.13	1.13[b]	1.14	1.14[b]
λ_{21}	1.16	0.75	1.07	1.07[b]	1.13	1.13[b]
λ_{31}	1.16	1.11	1.18	1.18[b]	1.12	1.12[b]
λ_{41}	0.90	0.83	0.92	0.92[b]	1.03	1.03[b]
λ_{51}	1.09	0.97	1.09	1.09[b]	0.97	0.97[b]
Ψ_{11}	1.00[a]	1.27	1.00[a]	1.02	1.00[a]	1.00
τ_{11}	4.65	4.65[b]	4.65	4.65[b]	4.61	4.61[b]
τ_{22}	4.80	4.37	4.80	4.64	4.75	4.75[b]
τ_{33}	5.18	5.56	5.18	5.58	5.27	5.27[b]
τ_{44}	5.07	4.59	5.07	4.64	4.86	4.86[b]
τ_{55}	4.35	5.01	4.35	5.08	4.58	4.58[b]
θ_{11}	0.94	0.78	0.93	0.84	0.91	0.83
θ_{22}	0.58	1.69	0.62	1.64	0.59	1.67
θ_{33}	0.49	0.69	0.48	0.73	0.53	0.82
θ_{44}	1.47	1.40	1.46	1.38	1.50	1.40
θ_{55}	1.84	2.02	1.82	2.00	1.93	2.15
α	−0.96		−0.92		−0.82	
χ^2	20.10		26.33		57.66*	
df	10		14		18	
χ^2 change	—		6.23		31.33*	
df change	—		4		4	
Tucker–Lewis index	0.96		0.97		0.91	
Goodness of fit	0.98		0.97		0.96	
Root mean square error of approximation	0.09		0.09		0.13	

Note: λ = item factor loadings. Ψ = factor variance. τ = item intercepts. θ = item residuals. α = latent mean difference between groups. [a]Fixed at 1.0. [b]Constrained to equality with first group in the same model. *$p < 0.01$.

ment invariance (i.e., equal intercepts) were not met. The observed mean differences on the items do not reflect the latent mean difference between the U.S. and Russian groups. Because the criteria for strong measurement invariance were not met, strict measurement invariance was not tested. Measurement invariance was next tested separately across the student participants and the community-based participants.

Comparing U.S. and Russian students

Next, measurement invariance for the SWLS was tested across U.S. and Russian student groups. As shown in Table IV, the baseline model resulted in a non-significant chi-square statistic and good practical fit values, indicating a one-factor model for both groups. Also, the results showed that the change in the chi-square statistic from the baseline model to the weak model was not significant, and this latter model showed good values of the practical fit indices. These results suggest that the relations between the items and the latent variable were similar across the U.S. and Russian student groups. Moreover, the strong measurement invariance model was not significantly different from the weak model (as evaluated by the change in chi-square values), and the strong model showed good practical fit indices, indicating that the group mean differences at the observed level reflected the group mean difference at the latent level. The criteria for strict measurement invariance were only marginally satisfied. There was a significant difference in model fit, as indicated by a decrease in the chi-square value from the strong to the strict measurement invariance models, but the values of the fit indices were not entirely unsatisfactory. Therefore, in addition to the items relating to the latent variable in the same way across groups, and the mean differences at the item level reflecting the mean difference at the latent level, the item residuals were approximately constant across groups. These results suggest that the SWLS is relatively free from measurement bias for U.S. and Russian student groups.

Comparing U.S. and Russian community members

Lastly, measurement invariance of the SWLS was tested for the U.S. and Russian community groups. As displayed in Table V, the results from the baseline model showed a significant chi-square value, as well

as poor values of the practical fit indices. Thus, a one-factor model for SWLS did not fit for both groups. Due to the lack of fit at this level, and because the baseline model was intended to serve as a basis of comparison with the more restrictive models, weak, strong, and strict measurement invariance were not tested. These findings indicate that SWLS scores are not comparable across the U.S. and Russian community-based groups.

TABLE IV

Parameter Estimates and Fit Indices for Testing the Baseline, Weak, Strong and Strict Measurement Invariance Models in the U.S. and Russian Student Samples

Parameter	Baseline		Weak		Strong		Strict	
	U.S.	Russia	U.S.	Russia	U.S.	Russia	U.S.	Russia
λ_{11}	0.94	0.94[b]	1.07	1.07[b]	1.05	1.05[b]	1.03	1.03[b]
λ_{21}	0.99	0.38	0.84	0.84[b]	0.86	0.86[b]	0.77	0.77[b]
λ_{31}	0.82	0.72	0.87	0.87[b]	0.83	0.83[b]	0.82	0.82[b]
λ_{41}	0.92	0.67	0.90	0.90[b]	0.96	0.96[b]	0.95	0.95[b]
λ_{51}	1.10	0.66	0.98	0.98[b]	1.03	1.03[b]	0.98	0.98[b]
Ψ_{11}	1.00[a]	2.15	1.00[a]	1.30	1.00[a]	1.27	1.00[a]	1.35
τ_{11}	4.61	4.61[b]	4.61	4.61[b]	4.66	4.66[b]	4.65	4.65[b]
τ_{22}	4.81	4.41	4.81	4.61	4.79	4.79[b]	4.75	4.75[b]
τ_{33}	5.36	5.51	5.36	5.54	5.43	5.43[b]	5.47	5.47[b]
τ_{44}	5.17	4.64	5.17	4.71	5.05	5.05[b]	5.03	5.03[b]
τ_{55}	5.64	5.19	5.64	5.31	5.56	5.56[b]	5.55	5.55[b]
θ_{11}	1.03	0.46	0.99	0.67	0.99	0.71	0.86	0.86[b]
θ_{22}	0.81	2.09	0.89	2.08	0.88	2.08	1.41	1.41[b]
θ_{33}	0.50	0.92	0.49	0.91	0.51	0.99	0.72	0.72[b]
θ_{44}	1.16	1.59	1.17	1.54	1.16	1.57	1.33	1.33[b]
θ_{55}	1.58	2.01	1.64	1.99	1.62	2.00	1.84	1.84[b]
α	−0.56		−0.50		−0.61		−0.62	
χ^2	5.16		15.03		23.30		41.41	
df	10		14		18		23	
χ^2 change	—		9.87		8.27		18.11*	
df change	—		4		4		5	
Tucker-Lewis index	1.04		0.99		0.97		0.92	
Goodness of fit	0.98		0.95		0.94		0.91	
Root mean square error of approximation	0.00		0.02		0.06		0.10	

Note: λ = item factor loadings. Ψ = factor variance. τ = item intercepts. θ = item residuals. α = latent mean difference between groups. [a]Fixed at 1.0. [b] Constrained to equality with first group in the same model. *$p < 0.01$.

TABLE V

Parameter Estimates and Fit Indices for Testing the Baseline Model in the U.S. and Russian Community Samples

Parameter	Baseline	
	U.S.	Russia
λ_{11}	1.22	1.22[b]
λ_{21}	1.38	0.95
λ_{31}	1.57	1.42
λ_{41}	0.85	0.79
λ_{51}	1.01	0.96
Ψ_{11}	1.00[a]	0.55
τ_{11}	4.71	4.71[b]
τ_{22}	4.79	4.29
τ_{33}	4.94	5.66
τ_{44}	4.94	4.31
τ_{55}	3.94	4.53
θ_{11}	0.82	0.83
θ_{22}	0.24	1.29
θ_{33}	0.22	0.56
θ_{44}	1.93	1.19
θ_{55}	2.05	1.83
α	−1.27	
χ^2	28.73*	
df	10	
χ^2 change	—	
df change	—	
Tucker–Lewis index	0.86	
Goodness of fit	0.92	
Root mean square error of approximation	0.17	

Note: λ = item factor loadings. Ψ = factor variance. τ = item intercepts. θ = item residuals. α = latent mean difference between groups. [a]Fixed at 1.0. [b]Constrained to equality with first group in the same model. *$p < 0.01$.

DISCUSSION

In this study, the SWLS was subjected to rigorous tests of the comparability of its scores across cultural groups. Criteria for weak measurement invariance were met when comparing U.S. and Russian samples. This finding supports conclusions from the literature regarding the existence of similar factor loadings for the SWLS across different groups (e.g., Pavot et al., 1991). A common practice among

researchers for checking for measurement bias at the weak level
involves observing the values of the factor loadings. The present
procedure, by contrast, provided a more formal process for evalu-
ating invariance of SWLS scores across two national groups.
Importantly, criteria for measurement invariance beyond the weak
level were not met. When comparing the U.S. and Russian groups,
the results showed a significant change in the chi-square value when
testing for strong measurement invariance, and poor values for the fit
indices, indicating that bias was introduced at the level of item means.
That is, the failure to meet criteria for strong measurement invariance
indicates that the item mean differences do not reflect latent mean
differences in the same way across groups. This finding suggests that
analyses testing for cultural differences in the SWLS should not be
conducted at the manifest level without concern that differences fail
to mirror latent variable differences.

   Due to expected age and cohort effects, measurement invariance
for the SWLS was tested separately in the student and in the
community-based samples. The results showed that criteria for weak,
strong, and perhaps strict measurement invariance were met when
comparing U.S. and Russian students. These findings suggest that,
at least for university students, the SWLS is moderately free from
measurement bias with respect to culture, measuring life satisfaction
similarly for students in two very different samples. Comparisons
of U.S. and Russian community participants, however, failed to
meet criteria for the least restrictive baseline model, indicating that a
one-factor model for the SWLS did not hold for both groups.

   The failure to meet criteria for the baseline model in testing for
measurement invariance across the older and more heterogeneous
community groups in the United States and Russia may be due to
differences in interpretations and conceptualizations of the items on
the SWLS. Specifically, differences in norms in expressing positive
affect, as well as in the value and desirability placed on attaining high
levels of life satisfaction, might have influenced responses to some of
the items, thus serving as a barrier to meeting criteria for measure-
ment invariance. However, because this proposition was not directly
tested, further investigation is needed into these and alternative
interpretations of our findings.

   Another potential source of the failure to meet criteria for
measurement invariance for the U.S. and Russian community-based

adults may be problems in the translation of the SWLS items. The meaning of some of the items, but not others, could have been altered during the translation of the original scale. (The reader is invited to compare the English and Russian versions of the SWLS [see Table I for the English version and Appendix A for the Russian version]). This potential source of bias can be ruled out, however, because the U.S. and Russian student samples met criteria for weak, strong, and strict measurement invariance. A third possibility is method bias (see van de Vijver and Leung, 1997; van de Vijver & Poortinga, 1997). Differences in response styles and familiarity with survey instruments may have led to the failure to meet criteria for the baseline model when comparing U.S. and Russian community groups. Russian community members were likely to have been relatively less experienced questionnaire consumers. Thus, they may have been overly concerned with the repetition inherent in the five SWLS items and may have tried to respond differentially to the nuances of each question.

Limitations and Conclusions

The most notable limitation of our study is the puzzling, yet to be explained, source of the problem in meeting criteria for the baseline model when comparing the U.S. and Russian community samples. It is still unclear whether the failure to establish a one-factor model (as tested in the baseline model) is due to the content of the items, or characteristics of our samples. The results merely suggest that the lack of fit for a one-factor model may have been due to differences in testing conditions, given that this is the most obvious difference among our samples; however, the possibility remains that other differences in the characteristics of the community samples could have contributed toward this failure (e.g., interpretations of certain words, or understanding of the intended meaning of one or some of the questions).

The importance of meeting criteria for weak, strong, or strict measurement invariance depends on the purposes for which the scale scores will be used. Researchers who are interested in group-differences may be most concerned with meeting criteria for strong measurement invariance. Meeting the additional criteria for strict measurement invariance is important for those who are interested in

correlations with external variables. Meeting criteria for strict measurement invariance is also important when one is interested in selection or placement decisions about individuals, to assure that the validity of the decisions are the same for all groups.[2]

It is a common practice among researchers to compare groups on scale scores without testing beyond the weak level of measurement invariance. In the present article, we argue that the consequence of not testing beyond the weak level (i.e., not testing for strong and strict measurement invariance) is severe – estimation of latent mean differences may be compromised. Moreover, comparisons of correlations with external variables across different groups may be imprecise, potentially leading researchers to faulty conclusions. Confirmation of measurement invariance permits meaningful group comparisons. Failure to confirm measurement invariance suggests that clear answers to substantive questions require improved measurement.

APPENDIX

The Russian Translation of the Satisfaction With Life Scale Figure 1.

NOTES

[1] This larger project was introduced to the participants as a study of personality, interests, and attitudes. Participants completed several questionnaires designed to assess their demographics, levels of happiness, satisfaction with life, self-esteem, optimism, personal goals and values, and other personal beliefs and attitudes. Each session lasted approximately one hour.

[2] These recommendations refer to the requirements for analyses with manifest variables (i.e., summated or averaged scores on observed variables). That is, researchers who use manifest level scale scores to study and make interpretations about latent constructs should meet criteria for strong measurement invariance when they are interested in mean differences, and strict measurement invariance when they are interested in comparing correlations across groups. When using structural equation modeling, however, meeting criteria for partial measurement invariance is sufficient.

Далее приводятся некоторые общие утверждения людей о жизни, о которой они имеют различные мнения и чувства.
Пожалуйста, укажите Ваше согласие или несогласие с этими утверждениями, обводя кружочком только один ответ для каждого утверждения.

1. В большинстве случаев моя жизнь близка к моему идеалу.

| Сильно не согласен | Не согласен | Немного не согласен | Нет согласия, нет также несогласия | Немного согласен | Согласен | Абсолютно согласен |

2. Условия моей жизни замечательны.

| Сильно не согласен | Не согласен | Немного не согласен | Нет согласия, нет также несогласия | Немного согласен | Согласен | Абсолютно согласен |

3. Я доволен моей жизнью.

| Сильно не согласен | Не согласен | Немного не согласен | Нет согласия, нет также несогласия | Немного согласен | Согласен | Абсолютно согласен |

4. До сих пор я всегда получал все важное в жизни, что я хотел.

| Сильно не согласен | Не согласен | Немного не согласен | Нет согласия, нет также несогласия | Немного согласен | Согласен | Абсолютно согласен |

5. Если бы я мог прожить мою жизнь еще раз, я бы почти ничего не изменил в ней.

| Сильно не согласен | Не согласен | Немного не согласен | Нет согласия, нет также несогласия | Немного согласен | Согласен | Абсолютно согласен |

Figure 1.

REFERENCES

Arrindell, W.A., J. Heesink and J.A. Feij: 1999, 'The Satisfaction With Life Scale (SWLS): Appraisal with 1700 healthy young adults in The Netherlands', Personality and Individual Differences 26, pp. 815–826.

Arrindell, W.A., L. Meeuwesen and F.J. Huyse: 1991, 'The Satisfaction With Life Scale (SWLS): Psychometric properties in a non-psychiatric medical outpatient sample', Personality and Individual Differences 12, pp. 117–123.

Atienza, F.L., I. Balaguer and M.L. García-Merita: 2003, 'Satisfaction With Life Scale: Analysis of factorial invariance across sexes', Personality and Individual Differences 35, pp. 1255–1260.

Balatsky, G. and E. Diener: 1993, 'Subjective well-being among Russian students', Social Indicators Research 28, pp. 225–243.

Bentler, P.M. and D.G. Bonett: 1980, 'Significance tests and goodness of fit in the analysis of covariance structures', Psychological Bulletin 88, pp. 588–606.

Blais, M.R., R.J. Vallerand, L.G. Pelletier and N.M. Brière: 1989, 'L'échelle de satisfaction de vie: Validation Canadienne-Francaise du "Satisfaction With Life Scale" [The satisfaction with life scale: Canadian-French Validation of the Satisfaction With Life Scale]', Canadian Journal of Behavioural Science 21, pp. 210–223.

Brislin, R.W.: 1970, 'Back-translation for cross-cultural research', Journal of Cross-Cultural Psychology 1, pp. 185–216.

Browne, M.W.: 1990, MUTMUM PC: User's guide (Ohio State University, Department of Psychology, Columbus).

Byrne, B.M., R.J. Shavelson and Muthén: 1989, 'Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance', Psychological Bulletin 3, pp. 456–466.

Carp, F.M. and A. Carp: 1983, 'Structural stability of well-being factors across age and gender, and development of scales of well-being unbiased for age and gender', Journal of Gerontology 38, pp. 572–581.

Diener, E. and M. Diener: 1995, 'Cross-cultural correlates of life satisfaction and self-esteem', Journal of Personality and Social Psychology 68, pp. 653–663.

Diener, E., R.A. Emmons, R.J. Larsen and S. Griffin: 1985, 'The satisfaction with life scale', Journal of Personality Assessment 49, pp. 71–75.

Diener, E., E.M. Suh, H. Smith and L. Shao: 1995, 'National differences in reported subjective well-being: why do they occur?' Social Indicators Research 34, pp. 7–32.

Drasgow, F.: 1984, 'Scrutinizing psychological tests: measurement equivalence and equivalent relations with external variables are the central issues', Psychological Bulletin 95, pp. 134–135.

Drasgow, F.: 1987, 'Study of the measurement bias of two standardized psychological tests', Journal of Applied Psychology 72, pp. 19–29.

Eid, M. and E. Diener: 2004, 'Global judgments of subjective well-being: situational variability and long-term stability', Social Indicators Research 65, pp. 245–277.

Freedman, J.: 1978, Happy People: What Happiness is, Who has it, and Why (Harcourt Brace Jovanovich, New York).

Grob, A., T. D. Little, B. Wanner, A. J. Wearing and Euronet: 1996, 'Adolescents' well-being and perceived control across 14 sociocultural contexts', Journal of Personality and Social Psychology, 71, pp. 785–795.

Jöreskog, K.G. and D. Sörbom: 1989, LISREL 7: A Guide to the Program and Applications, 2nd edition (SPSS, Chicago).

Liang, J. and K.A. Bollen: 1985, 'Gender differences in the structure of the Philadelphia Geriatric Center Morale Scale', Journal of Gerontology 40, pp. 468–477.

Liang, J., R.H. Lawrence and K.A. Bollen: 1986, 'Age differences in the structure of the Philadelphia Geriatric Center Morale Scale', Psychology and Aging 1, pp. 27–33.

Liang, J., R.H. Lawrence and K.A. Bollen: 1987, 'Race differences in factorial structures of two measures of subjective well-being', Journal of Gerontology 42, pp. 426–428.

Lyubomirsky, S.: 1997, The Meaning and Expression of Happiness: Comparing the United States and Russia. Paper presented at the Ninth Conference of the American Psychological Society, Washington DC.

Meredith, W.: 1993, 'Measurement invariance, factor analysis and factorial invariance', Psychometrika 58, pp. 525–543.

Oishi, S., E.F. Diener, R.E. Lucas and E.M. Suh.: 1999, 'Cross-cultural variations in predictors of life satisfaction: perspectives from needs and values', Personality and Social Psychology Bulletin 25, pp. 980–990.

Pavot, W. and E. Diener: 1993, 'Review of the Satisfaction With Life Scale', Psychological Assessment 5, pp. 164–172.

Pavot, W., E. Diener, C.R. Colvin and E. Sandvik: 1991, 'Further validation of the Satisfaction With Life Scale: evidence for the cross-method convergence of well-being measures', Social Indicators Research 28, pp. 1–20.

Reise, S., K.F. Widaman and R.H. Pugh: 1993, 'Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance', Psychological Bulletin 114, pp. 552–566.

Ryan, R.M., V.I. Chirkov, T.D. Little, K.M. Sheldon, E. Timoshina and E.L. Deci: 1999, 'The American dream in Russia: extrinsic aspirations and well-being in two cultures', Personality and Social Psychology Bulletin 25, pp. 1509–1524.

Sam, D.L.: 2001, 'Satisfaction with life among international students: An exploratory study', Social Indicators Research 53, pp. 315–337.

Schimmack, U., P. Radhakrishnan, S. Oishi, V. Dzokoto and S. Ahadi: 2002, 'Culture, personality, and subjective well-being: integrating process models of life satisfaction', Journal of Personality and Social Psychology 82, pp. 582–593.

Shevlin, M.E., V. Brunsden and J.N.V. Miles: 1998, 'Satisfaction With Life Scale: Analysis of factorial invariance, mean structures, and reliability', Personality and Individual Differences 25, pp. 911–916.

Shevlin, M.E. and B.P. Bunting: 1994, 'Confirmatory factor analysis of the Satisfaction With Life Scale', Perceptual and Motor Skills 79, pp. 1316–1318.

Smith, H.: 1990, The New Russians (Random House, New York).

Steiger, J.H. and J. Lind: 1980, Statistically based Tests for the Number of Common Factors. Paper presented at the meeting of the Psychometric Society, Iowa City, IA.

Tanaka, J.S. and G.J. Huba: 1985, 'A fit index for covariance structure models under arbitrary GLS estimation', British Journal of Mathematical and Statistical Psychology 38, pp. 197–201.

Triandis, H.C., R. Bontempo, K. Leung and C.H. Hui: 1990, 'A method for determining cultural, demographic, and personal constructs', Journal of Cross-Cultural Psychology 21, pp. 302–318.

Tucker, L.R. and C. Lewis: 1973, 'A reliability coefficient for maximum likelihood factor analysis', Psychometrika 38, pp. 1–10.

van de Vijver, F.J.R. and K. Leung: 1997, 'Methods and data analysis of comparative research', in J.W. Berry and Y.H. Poortinga (eds.), Handbook of Cross-cultural Psychology, Vol. 1, Theory and Method (Allyn & Bacon, Inc, Boston), pp. 257–300.

van de Vijver, F.J.R. and Y.H. Poortinga: 1997, 'Towards an integrated analysis of bias in cross-cultural assessment', European Journal of Psychological Assessment 13, pp. 29–37.

Veenhoven, R.: 1991, 'Is happiness relative?', Social Indicators Research 24, pp. 1–34.

Veenhoven, R.: 1996, 'Developments in satisfaction research', Social Indicators Research 37, pp. 1–46.

Widaman, K.F. and S.P. Reise: 1997, 'Exploring the measurement invariance of psychological instruments: applications in the substance use domain', in K. J. Bryant, M. Windle and S.G. West (eds.), The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research (American Psychological Association, Washington DC).

Wierzbicka, A.: 1994, 'Emotion, language, and cultural scripts', in S. Kitayama and H.R. Markus (eds.), Emotion and Culture: Empirical Studies of Mutual Influences (American Psychological Association, Washington DC), pp. 133–196.

Department of Psychology   Kari L. Tucker
Irvine Valley College
5500 Irvine Center Drive
Irvine, CA 92618
USA

Department of Psychology   Daniel J. Ozer
University of California   Sonja Lyubomirsky
Riverside, CA 92521   Julia K. Boehm
USA
E-mail: sonja.lyubomirsky@ucr.edu