Slovak validation of the Levels of Self-Criticism Scale: An Item Response Theory analysis

Julia Halamová[1*], Martin Kanovský[2] and Monika Pacúchová[1]

[1]Institute of Applied Psychology, Faculty of Social and Economic Sciences, Comenius University in Bratislava, Slovakia
[2]Institute of Social Anthropology, Faculty of Social and Economic Sciences, Comenius University in Bratislava, Slovakia

Abstract: Severe self-criticism is one of the underlying psychological factors which most influence susceptibility to psychopathology and its persistence. This study verifies the psychometric features and factor structure of the Slovak version of the Levels of Self-Criticism Scale. It entails the use of descriptive analysis, reliability analysis, validity analysis, and IRT confirmatory factor analysis. As opposed to the English version with two dimensions (Internalized Self-Criticism and Comparative Self-Criticism) using classical factor analysis, the Slovak translation has three factors (Internalized Self-Criticism, Unfavourable Comparison with Others and Favourable Comparison with Others) exploiting more advanced IRT factor analysis. It is important to explore this question in more detail in future, especially because the proposed three-dimensional structure is not very meaningful from the purely theoretical perspective. The Levels of Self-Criticism Scale appears to have acceptable level of reliability and validity. Therefore, we recommend using this instrument to assess the levels of self-criticism in the English and Slovak versions.

Keywords: Levels of Self-Criticism Scale, self-criticism, test construction, Item Response Theory analysis

Slovaška validacija Lestvice stopenj samokritičnosti: analiza s teorijo odgovora na postavko

Julia Halamová[1*], Martin Kanovský[2] in Monika Pacúchová[1]

[1]Inštitut uporabne psihologije, Fakulteta za socialne in ekonomske znanosti, Comenius univerza v Bratislavi, Slovaška
[2]Inštitut socialne antropologije, Fakulteta za socialne in ekonomske znanosti, Comenius univerza v Bratislavi, Slovaška

Povzetek: Izrazita samokritičnost spada med najpomembnejše psihološke dejavnike dovzetnosti za nastanek in vztrajanje psihopatologije. V raziskavi smo preverili psihometrične značilnosti in faktorsko strukturo slovaške različice Lestvice stopenj samokritičnosti. Uporabili smo opisno analizo, analizo zanesljivosti in veljavnosti ter konfirmatorno faktorsko analizo v okviru teorije odgovora na postavko (TOP). V nasprotju z angleško različico, kjer je klasična faktorska analiza pokazala dve dimenziji (ponotranjeno samokritičnost in primerjalno samokritičnost), je pri slovaški različici bolj sofisticirana faktorska analiza v okviru TOP pokazala obstoj treh faktorjev (ponotranjena samokritičnost, neugodna primerjava z drugimi in ugodna primerjava z drugimi). V prihodnje bo treba ta problem še podrobneje raziskati, ker predlagana trifaktorska struktura s povsem teoretičnega vidika ni najbolj smiselna. Zdi se, da ima Lestvica ravni samokritičnosti sprejemljivo raven zanesljivosti in veljavnosti, zato lahko priporočamo njeno uporabo za oceno ravni samokritičnosti z angleško in slovaško različico.

Ključne besede: Lestvica stopenj samokritičnosti, samokritičnost, sestava testa, teorija odgovora na postavko

* Naslov/Address: Julia Halamová, PhD, Institute of Applied Psychology, Faculty of Social and Economic Sciences, Comenius University in Bratislava, Mlynske luhy 4, 821 05 Bratislava, Slovakia, e-mail: julia.halamova@gmail.com

There are several scales that measure self-criticism. One of the first measures was the Depressive Experiences Questionnaire (DEQ; Blatt, D'Afflitti, & Quinlan, 1979). Related to this instrument is the Levels of Self-Criticism Scale (LOSC; Thompson & Zuroff, 2004). The Forms of Self-Criticizing / Attacking & Self-Reassuring Scale (FSCRS; Gilbert, Clark, Hempel, Miles, & Irons, 2004) was also created. The LOSC and FSCRS are the only scales that treat self-criticism as a multi-dimensional construct. The only scale that assesses situational self-criticism is the Self-Compassion and Self-Criticism Scales (SCCS; Falconer, King, & Brewin, 2015). Thanks to our research grant we translated all of these scales and analysed their psychometric properties and factor structures in order to conduct further research on self-criticism in Slovakia. Having a valid and reliable measurement instrument is an essential precondition for self-criticism research. It appears that apart from English, Slovak is the only language into which all of the scales have been translated and psychometrically analysed (Halamová & Kanovský, 2017). So far, psychometric analysis of the following scales in Slovak language were published: SCCS (Halamová, Kanovský, & Pacúchová, 2018), SCS (Halamová, Kanovský, & Pacúchová, 2018), and FSCRS (Halamová, Kanovský, & Pacúchová, 2017). Therefore, in this study we focused on the psychometric analysis of Slovak version of the LOSC.

Up to now, most of the research studies on self-criticism have been grounded on Sydney Blatt's theory (1974). Besides, the LOSC was created on the ground of this theory. Blatt (1974) sees self-criticism as an abnormal and dysfunctional introjective development which places individuals at risk of psychopathology. There are two variations of self-criticism according to Thompson and Zuroff (2004), whose research is also based on Blatt's theory (1974). One form of self-criticism is constructed on the externalized criteria of dominant others. These externalized standards are perceived as animosity and opprobrium from people, and a sense of inferiority when compared with them (Thompson & Zuroff, 2004). Individuals who possess this variation in self-criticism avoid dealing with problems which they see as insurmountable. This is called Comparative self-criticism (CSC; Thompson & Zuroff, 2004, p. 419) and is defined as "*the unfavourable comparison with others, who are seen as superior and as hostile or critical; consequently, there is discomfort with being evaluated or exposed to others.*" Individuals with this form of self-criticism do not trust others and try to avoid or emulate them. In contrast, Internalized self-criticism (ISC; Thompson & Zuroff, 2004, p. 420) is described as the "*negative view of the self in comparison with internal, personal standards. These internal standards tend to be both high and constantly receding, resulting in a chronic failure to meet one's own standards.*" Therefore, people with this kind of self-criticism have a chronic sense of inferiority of one's own impossibly high standards. Extraordinary standards are not by themselves indicative of self-criticism. It occurs even when these extraordinary personal standards are met, the person is not satisfied, and does not prize or recognize their own success. Instead, the person with high self-criticism reacts to the accomplishment by increasing the standards for success and therefore is never satisfied with any achievement no matter how great.

The development of the Levels of Self-Criticism Scale

The Levels of Self-Criticism Scale (LOSC; Thompson & Zuroff, 2004) consists of two subscales: Comparative Self-Criticism (CSC) and Internalized Self-Criticism (ISC). The first version of the scale contained 34 items. Each item was designed to describe one of the forms of self-criticism and to discriminate between them. Participants indicate the level of agreement with a certain item on a Likert type scale (ranging from 1 = *not at all*, to 7 = *very well*). After conducting factor analysis, the items in the scale were reduced to 22, with the CSC factor (12 items) and the ISC factor (10 items). The items were reduced according to the following criteria (Thompson & Zuroff, 2004, p. 422): "*Items that did not load on either factor (above .25), or that loaded highly (above 0.5) on both factors were dropped from the scale. As well, items that correlated highly with the BIDR [Balanced Inventory of Desirable Responses] (above .3) and items that did not have an adequate variance of responses (standard deviation less than 1) were dropped from the scale*". A higher total score on the entire scale indicates a higher level of self-criticism (Thompson & Zuroff, 2004).

Basic psychometric properties of the LOSC Scale

The psychometric features of the scale were explored in the original publication by Thompson and Zuroff (2004). The participants were undergraduate students from the US. The original factor analysis confirmed the two-factor solution of the LOSC. With respect to the expected correlation between both factors, the authors carried out a factor analysis (not specified) with oblique rotation (Thompson & Zuroff, 2004). The correlation between the two ISC and CSC subscales was moderately high ($r = .45$, $p < .05$). The Comparing Self-Criticism factor has 10 items, such as "I fear that if people get to know me too well, they will not respect me." The factor loadings of the subscale ranged from a low of .31 to a high of .72. The Internalized Self-Criticism factor has 12 items, such as "I am very irritable when I have failed." and the factor loadings of the subscale ranged from .51 to .89 (Thompson & Zuroff, 2004).

Although many research studies use the English language version of the LOSC (e.g., Smart, Peters, & Baer, 2016), there are translations into various languages (Chinese: Chen et al., 2016; Italian: Manfredi et al., 2016; Japanese: Yamaguchi & Kim, 2013; Persian: Ghorbani, Mousavi, Watson, & Chen, 2011; Portuguese: Castilho, Pinto-Gouveia, & Duarte, 2015; Romanian: Crăciun, 2013; Slovak: Halamová, Kanovský, & Pacúchová, 2017, 2018; Swedish: Danielsson, 2012; and Turkish: Öngen, 2006). However, to date, no other research study than the original study (Thompson & Zuroff, 2004) has verified the LOSC factor structure. Also, the original study by Thompson and Zuroff used factor analysis with oblique rotation but not more advanced forms of statistics such as item response theory factor analysis. IRT methods are more appropriate if the items are not continuous: since the LOSC

items are ordinal because of Likert scoring, IRT analysis with a logistical approach outperforms the standard analysis with a linear approach.

The authors (Thompson & Zuroff, 2004) found a good level of Cronbach's alpha for the LOSC subscales in two subsequent research studies (CSC α = .81 / .84 ISC α = .87 / .88). Yamaguchi, Kim and Akutsu (2014) reported an acceptable level of Cronbach's alpha (.78 for the CSC and .84 for ISC) among Japanese college students and US undergraduate students. It is not explicitly reported in the article whether the scale was translated into Japanese language or not. In the research study by Smart et al. (2016) with undergraduate students from the US, the ISC subscale achieved excellent internal consistency (α = .92) and the CSC subscale achieved a good level of internal consistency (α = .80).

Furthermore, the authors tested the construct validity of the scale (Thompson & Zuroff, 2004). Verifying construct validity was based on an assessment of the correlation between CSC and ISC with constructs such as self-criticism, self-esteem and perfectionism, all of which should be theoretically interrelated. The authors confirmed the correlation between CSC, ISC and the Self-Criticism subscale from the Depressive Experiences Questionnaire DEQ (Blatt, D'Afflitti, & Quinlan, 1976) ranging from .55 to .62 (Thompson & Zuroff, 2004), and indicating positive correlations with perfectionism (from .21 to .49 measured by the Multidimensional Perfectionism Scale; Hewitt & Flett, 1989) and negative correlations with self-esteem (from .52 to .66 measured by the Rosenberg Self-Esteem Scale; Rosenberg, 1965). Compared to the CSC subscale, the ISC was more strongly correlated with Self-Oriented Perfectionism. On the contrary, the CSC was more highly correlated with self-esteem and with DEQ self-criticism (Thompson & Zuroff, 2004).

Validity was also explored via relationships between the ISC and CSC with personality constructs of the Big Five (measured by The NEO Five-Factor Inventory; Costa & Mc-Crae, 1992). The CSC and ISC subscales both showed strong positive correlations with Neuroticism (.60 and .54, respectively), but only CSC had negative correlations with Extraversion (–.37), Conscientiousness (–.34) and Agreeability (–.35). According to the authors (Thompson & Zuroff, 2004), this result supports the distinctiveness of these two factors of the LOSC.

The construct validity was determined by measuring the relationships of the CSC and ISC with the style of attachment (measured by the Attachment Scales; Bartholomew & Horowitz, 1991) and the interpersonal conflicts style (measured by the Conflict Styles Inventory; Levinger & Pietromonaco, 1989). Positive correlations were confirmed between the CSC and ISC and preoccupied (.47 and .25) and the avoidant (.30 and .18) attachment styles. According to Bowlby (1980) and Bartholomew and Horowitz (1991), there are four attachment styles: secure (with low avoidance and anxiety in close relationships), preoccupied (with low avoidance but high anxiety), avoidant (with high avoidance as well as anxiety), and dismissing (with high avoidance but low anxiety).

Negative correlations were shown between the CSC and ISC and the secure attachment style (–.44 and –.18). No relationship with a dismissing style of attachment was found. Interestingly, the correlations were stronger for CSC than for ISC supporting the distinctiveness of the two factors. The LOSC was positively correlated with avoiding and accommodating conflict styles, and negatively correlated with compromising and contending conflict styles (Thompson & Zuroff, 2004). Levinger and Pietromonaco (1989) described how people deal with conflicts in close relationships and distinguish between avoiding style, which means that people prefer avoiding conflicts, collaborating style, which means working together to find mutual satisfaction, compromising style, which means finding solution half way between people's needs, accommodating style, which means conforming to the partner's needs and contending style, which means seeking your own needs first.

Thompson and Zuroff (2004) selected self-criticism from the DEQ as a validity criterion because they wanted to evaluate convergent validity by the same construct but using a different instrument. Since then, a new self-criticism measuring instrument, which is the FSCRS, and a new self-compassion measuring instrument, which is the Self-Compassion Scale, were developed (Neff, 2003a). Self-compassion (Gilbert & Choden, 2013) is defined as sensitivity to the suffering of self and other people accompanied by commitment to try to relieve and avert suffering. On the top of that, as Neff and Dahm (2015, p. 122) stated: *"... a key feature of self-compassion is the lack of self-criticism ..."* Therefore, self-compassion is considered to be a crucial antidote for self-criticism (Gilbert & Irons, 2009; Greenberg, 2011; Neff, 2003b) and could be used as a discriminant validity criterion for self-criticism. To the best of our knowledge, no other study than the original one (Thompson & Zuroff, 2004) has investigated the validity of the LOSC scale. Hence, it is important to evaluate the validity of the Slovak version of the LOSC by other instruments such as the FSCRS and the SCS.

Aim of the study

The purpose of this paper was to investigate the psychometric properties and factor structure of the Slovak version of the Levels of Self-Criticism Scale through item response theory (IRT) and to verify whether it corresponds to the original factor structure (Thompson & Zuroff, 2004). We also estimated the reliability of the scale and explored its validity by correlations with the scales measuring self-criticism and self-compassion. Moreover, we expected that the LOSC subscales would be correlated significantly with self-criticism (Hated and Inadequate Self of the FSCRS, Self-Uncompassionate Responding of the SCS – positive correlations), and self-compassion (Self-Compassionate Responding of the SCS – negative correlations), while the levels of reliability of the LOSC subscales and factor structure of the scale would be similar to the results attained in previous LOSC studies.

Method

Participants

The research sample included 592 participants, of which 264 were males (45%) and 328 were females (55%). The participants ranged in age from 18 to 82 years and the mean age was 34.96 years *(SD* = 14.17). Concerning highest level of education, 18.9% of the participants were elementary school graduates, 47.8% were high school graduates, and 33.3% were university graduates. 51% of the participants were single and 49% were married.

Instruments

The Levels of Self-Criticism Scale (LOSC). The LOSC (Thompson & Zuroff, 2004) is a 22-item instrument with two dimensions: the Comparative Self-Criticism dimension (CSC) and the Internalized Self-Criticism dimension (ISC). The first author of the study together with her research team used an expert panel to do the forward translation and then a professional independent translator did a back translation of the scale. Afterwards, the discrepancies were discussed and approved by consensus by the co-authors and the translator to ensure that the Slovak version bore a resemblance to the English version of the LOSC.

The Forms of Self-Criticizing / Attacking & Self-Reassuring Scale (FSCRS). The FSCRS (Gilbert et al., 2004) is a 22-item instrument with three dimensions: the Reassured Self (RS), Inadequate Self (IS), and Hated Self (HS). The scale was developed to measure self-criticism in two forms (the IS focuses on feelings of personal inadequacy and the HS measures the desire to hurt or punish oneself) and the ability to self-reassure. This scale measures different ways in which people think and feel about themselves when something goes wrong in their lives. Participants were asked to indicate the degree to which they agree with the particular statement on a 5-point Likert type scale (ranging from 0 = *not at all like me* to 4 = *extremely like me*). The FSCRS was translated into Slovak by Halamová, Kanovský and Pacúchová (2017). In a Slovak convenience sample of 1181 participants, the three-dimensional structure of the FSCRS was confirmed (Halamová et al., 2017) and the reliability coefficients varied between .74 and .92 for the entire scale and for the subscales.

The Self-Compassion Scale (SCS). The SCS (Neff, 2003a) is a 26-item instrument which measures six aspects/ subscales of self-compassion in situations of perceived difficulty: Self-Kindness (SK), Self-Judgement (SJ), Common Humanity (CH), Isolation (IS), Mindfulness (MI) and Over-Identification (OI). Participants were asked to indicate the degree to which they agree with the particular statement on a 5-point Likert type scale (ranging from 1 = *almost never* to 5 = *almost always*). Self-Kindness represents the ability to take care of oneself and be warm towards oneself when encountering failure situations. The Common Humanity subscale reflects the personal understanding that suffering is part of the shared human experience. Mindfulness is a non-judgmental state of mind in which individuals observe their thoughts and feelings as they are, without over-identification, suppression or denial. The scale measures the degree to which individu-

als display self-kindness against self-judgment, common humanity against isolation, and mindfulness against over-identification. The SCS was translated into Slovak language by Halamová, Kanovský and Pacúchová (2018). In the Slovak convenience samples of 1857 participants, a six-dimensional model or one higher order factor of SCS were not confirmed (Halamová et al., 2018); therefore, the use of total score or six-dimensions is inappropriate. Instead, two general factors (Self-Compassionate Responding and Self-Uncompassionate Responding) were identified. These two factors represent positive and negative items grouped separately. However, the scale showed acceptable reliability properties with coefficients between .68 and .92 for the two higher–order factors as well as the six factors.

Procedure

The data was collected by convenience and snowball sampling as part of the research grant focused on self-criticism and self-compassion. Questionnaires were distributed in digital form via social networks to university students and their friends and family members. The questionnaire was designed to ensure that all of the questions were answered in order. Participants were given informed consent forms together with the scales and basic demographic questions. The authors have no conflicts of interest. The research study was performed in compliance with the Code of Ethics of the corresponding university and APA ethical principles.

Data analysis

The following software was used in this study: R (version 3. 1. 3., R Core Team, 2015), libraries psych (Revelle, 2015) and mirt (Chalmers, 2012). The analysis procedure was as follows:

1. Descriptive analysis: standard distribution properties of LOSC items, as well as testing of the univariate normal distribution of the LOSC. Due to the ordinal nature of the data we did not assume a normal distribution: Likert-type items are ordinal by nature, and the thesis that they are approximately interval is indefensible (Bishop & Herron, 2015). Skewness was tested by the D'Agostino test (D'Agostino, 1970); testing the null hypothesis that skewness equals zero. Kurtosis was tested by the Ansombe-Glynn test (Anscombe & Glynn, 1983) testing the null hypothesis that kurtosis equals 3. The normal distribution of items was tested by the robust Jarque-Bera test. Classical tests of normality are sensitive to outliers, but the robust Jarque-Bera test uses a robust standard deviation (the average absolute deviation from the median), therefore it is not biased by outliers (Gel & Gastwirth, 2008). According to the null hypothesis, the data followed a normal distribution. Due to the polytomous ordinal character of items, we used the polychoric item-total correlation.

2. IRT confirmatory factor analysis. If items are polytomous ordinal, IRT models are generally more appropriate: the full-information logistic models outperform linear models in their method of estimation (Maydeu-Olivares,

Cai & Hernandez, 2011). This is due to the fact that since IRT methods estimate several thresholds per item, and not only a single intercept, they take into account an ordinal character of items. Estimating single intercepts (means) is not advisable, especially when items are heavily skewed. We agreed that 7-point Likert-type items could be analysed by linear methods as well (e.g., a robust maximum likelihood estimator, or a weighted least squares means and variances adjusted estimator with thresholds; Li, 2016; Rhemtulla, Brosseau-Liard, & Savalei, 2012), but since these estimators use limited-information estimation, the full-information estimation is more appropriate. Multidimensional IRT methods are already widely used in psychological research and are available in many accessible statistical software programs. We tested the fit of two-factor, three-factor and bi-factor IRT models. The two-factor model would replicate the original factor structure of LOSC. The three-factor model would test the notion that one of the original factors is in fact composed of two factors, as suggested in parallel analysis (Dinno, 2009). And finally, the bi-factor model would test the assumption that despite multidimensionality, one strong general factor could account for a substantial degree of explained variance. Unlike the higher-order model, the bi-factor model allows for the decomposition of the explained variance directly into a single general factor, particular residual factors, and errors. The higher-order model only unifies variance explained by particular factors, but does not enable its decomposition. Metaanalysis of 166 comparisons showed (Cucina & Byle, 2017) that the bi-factor model fits better than the higher-order model in more than 90% of comparisons. Unlike the higher-order model, the bi-factor model allows for calculating the reliability index (the hierarchical Omega) which directly informs on proportion of explained variance by the single general factor. For the verification of the conformity of the model with the data, we used standard fit indices (CFI – Comparative Fit Index, TLI – Tucker-Lewis Index, RMSEA – Root Mean Squared Error of Approximation, SRMR – Standardized Root Mean Residuals), for which the following recommended thresholds are commonly used in literature: CFI and TLI > .90 (acceptable fit; Bentler, 1992) or > .95 (excellent fit; Hu & Bentler, 1999), RMSEA < .08 (acceptable fit) or < .05 (excellent fit; Browne & Cudeck, 1993), SRMR < .08 (acceptable fit) or < .05 (excellent fit; Hu & Bentler, 1999). Horn's parallel analysis (Horn, 1965; Dinno, 2009) was conducted to justify the appropriate number of factors. Likelihood-ratio tests were used to compare nested models, and we also report the information criteria (AIC and BIC).

3. Analysis of the reliability of each dimension of the instrument. The Cronbach α index, the most commonly used index of reliability, can be very inaccurate when used for ordinal scales (Zumbo, Gadermann, & Zeisser, 2007; Dunn, Baguley, & Brunsden, 2014). This uncertainty can be partially corrected if it is not calculated from the Pearson's correlation matrix but from the polychoric correlation matrix, which takes into account the ordinal nature of the variables (Zumbo et al., 2007). Therefore, for the analysis of reliability we used the Cronbach α index calculated from the polychoric correlation matrix.

4. Exploration of validity by means of Spearman correlation coefficients with the FSCRS and SCS scales and their dimensions. Spearman correlations are used because the scales of raw scores failed to display normal distribution. In the case of the FSCRS, it was related to self-criticism in the form of Hated and Inadequate Self and Reassured Self as the opposite construct. In the case of the SCS, we tested the relationship with Self-Uncompassionate responding which is the negative, opposite part of self-compassion closely reminiscent of self-Criticism and Self-Compassionate responding which should have a negative relationship to self-criticism.

Results

Descriptive analysis

A descriptive statistical analysis of LOSC items can be found in Table 1. Since items are ordinal, their non-normal distribution was assumed, which was confirmed: All items displayed significant kurtosis. The values of kurtosis were not normalized by substracting 3 (excess kurtosis): therefore the value of the kurtosis for a standard normal distribution is 3 and not 0. The values of kurtosis are rather low (platykurtic). Many of them are heavily skewed (five had positive skew, three had negative skew) and all of them follow a non-normal distribution according to robust Jarque-Bera test. Descriptive data showed that items 6, 11, and 22 had undesirable psychometric properties (their correlation to the total score was below .30), and were discarded in the following analyses.

Factor analysis

As already stated, there is no hope that the ordinal variables that make up the items of the questionnaire could meet the assumption of multivariate normal distribution, which is essential for the correct functioning of the classical factor analysis (based on the maximum likelihood method). Therefore, for factor analysis, the IRT method (item-response theory), which is much more relevant and accurate for analysing ordinal variables given the logistic and not the linear method of its estimation, was used. For factor analysis, the Samejima confirmation model (graded response model) was used in the "mirt" package (Samejima, 1969), namely its multidimensional extension (Samejima, 1974; see also McDonald, 1997). Full-information estimation with Metropolis-Hastings Robbins-Monro algorithm was used (Cai, 2010).

The original 2-dimensional IRT model (Samejima graded-response model) did not demonstrate a good fit to the data (CFI = .70, TLI = .60, RMSEA = .12 (90 % CI = .11–.13, SRMR = .12, AIC = 39032, BIC = 39619). Since this estimation method is full-information, χ^2 statistics based on covariance matrix are not available. Parallel analysis (Dinno, 2009) showed that the dimension of "Comparative Self-Criticism" in the Slovak version is composed of two dimensions: items 2, 4, 10, 14, 18 are grouped in the sub-dimension named Unfavourable Comparison with Others (USC), while items 8, 12, 16, 20, 21 are grouped in the second sub-dimension named Favourable Comparison with Others (FSC). The values of the fit indices for the three-factor IRT Samejima graded response

Table 1. *Descriptive statistics and distribution properties of LOSC items*

LOSC item	*M*	*SD*	*Skewness*	*Kurtosis*	*RJB*	*ISC*
LOSC-1	4.38	1.60	−0.19	2.14***	18.28***	.52
LOSC-2	3.37	1.70	0.40***	2.22***	22.20***	.65
LOSC-3	4.30	1.63	−0.31**	2.40***	18.73***	.55
LOSC-4	4.23	1.72	−0.09	2.02***	20.51***	.44
LOSC-5	4.19	1.61	−0.24*	2.23***	15.32***	.62
LOSC-6	3.76	1.74	0.08	2.06***	14.93***	.11
LOSC-7	3.88	1.62	−0.07	2.20***	7.04*	.55
LOSC-8	4.23	1.54	−0.17	2.44***	6.52*	.36
LOSC-9	3.90	1.58	0.10	2.20***	10.96**	.58
LOSC-10	3.10	1.76	0.62***	2.43***	40.44***	.57
LOSC-11	3.97	1.65	0.05	2.14***	11.03**	.29
LOSC-12	4.71	1.85	−0.42***	2.05***	28.97***	.32
LOSC-13	3.62	1.73	0.16	2.08***	19.37***	.59
LOSC-14	4.18	1.67	0.01	2.10***	14.10***	.35
LOSC-15	4.15	1.57	−0.10	2.15***	14.22***	.56
LOSC-16	4.15	1.61	−0.03	2.21***	6.81*	.32
LOSC-17	3.52	1.50	0.17	2.35***	13.00**	.55
LOSC-18	2.85	1.63	0.66***	2.49***	41.92***	.52
LOSC-19	4.07	1.70	−0.22*	2.22***	11.43**	.44
LOSC-20	4.11	1.76	−0.06	2.02***	15.09***	.35
LOSC-21	4.68	1.71	−0.23*	2.06***	21.20***	.45
LOSC-22	4.05	1.76	0.03	1.94***	19.05***	.15

Notes. RJB = Robust Jarque-Bera test of normal distribution. *ISC* = corrected polychoric correlation of item to total score. LOSC - The Levels of Self-Criticism Scale.
N = 592. * *p* < .05, ** *p* < .01, *** *p* < .001.

Table 2. *Factor loadings of confirmatory IRT (Samejima graded response) models of LOSC*

LOSC item	ISC	USC	FSC	*h²*
LOSC-1	.706	.000	.000	.499
LOSC-2	.000	.793	.000	.629
LOSC-3	.720	.000	.000	.518
LOSC-4	.000	.551	.000	.304
LOSC-5	.806	.000	.000	.650
LOSC-7	.669	.000	.000	.448
LOSC-8	.000	.000	.584	.342
LOSC-9	.700	.000	.000	.490
LOSC-10	.000	.711	.000	.506
LOSC-12	.000	.000	.792	.628
LOSC-13	.710	.000	.000	.504
LOSC-14	.000	.290	.000	.084
LOSC-15	.718	.000	.000	.515
LOSC-16	.000	.000	.551	.304
LOSC-17	.656	.000	.000	.431
LOSC-18	.000	.665	.000	.442
LOSC-19	.562	.000	.000	.316
LOSC-20	.000	.000	.538	.290
LOSC-21	.000	.000	.714	.510

Notes. h² = explained variance. LOSC = The Levels of Self-Criticism Scale. LOSC ISC = Internalized self-criticism. LOSC USC = Unfavourable comparison with others. LOSC FSC = Favourable comparison with others.
N = 592

model were as follows: CFI = .94, TLI = .92, RMSEA = .051 (90% confidence interval .041–.062), SRMR = .079, AIC = 38518, BIC = 39114. A likelihood-ratio test (and lower information criteria) showed that the 3-factor model had a better fit than the 2-factor model, $\chi^2(2) = 518$, $p < .001$. Factor loadings and explained variances are shown in Table 2. All factor loadings have acceptable values (>.30) except for the factor loading of item number 14 which had a value of .29 and is therefore least acceptable.

All of the fit indices for this model are very good or at least acceptable. The correlation of the latent factors was .82 between the ISC and USC, .01 between the ISC and FSC and .26 between the USC and FSC. The values of the explained variance (sums of squared loadings / number of items) are 23% for ISC, 10% for USC and 11% for FSC. The IRT parameters and their standard errors (computed by bootstrap procedure with 100 draws) are shown in Table 3. The average standard error is 0.18, so despite the large number of parameters and moderate sample size, the parameters have been estimated with an appropriate precision.

Item information curves for the dimension Internalized Self-Criticism (Figure 1A) show that item 5 has the larger information, and item 19 the lowest information. Test information curve (Figure 1B) shows that this dimension measures better the lower parts of latent ability (self-criticism). Item information curves for the dimension Unfavorable Comparison with Others (Figure 1C) show that items 2, 10 and 18 have large information, item 4 low information, and item 14 contributes practically nothing to test information (as expected

Table 3. *IRT parameters and their standard errors of the 3-factor comfirmatory model for LOSC*

Dimension		Item	Slope(SE)	Thresholds(SE)
Internalized Self-Criticism – ISC		ISC1	1.70(0.13)	4.64(0.35), 2.54(0.20), 1.09(0.15), 0.09(0.13), –1.41(0.15), –3.34(0.22)
		ISC3	1.77(0.16)	4.05(0.29), 2.19(0.17), 1.12(0.15), –0.03(0.13), –1.59(0.16), –3.64(0.25)
		ISC5	2.32(0.19)	4.47(0.32), 2.54(0.21), 1.04(0.17), –0.28(0.15), –2.08(0.20), –4.50(0.30)
		ISC7	1.53(0.15)	3.23(0.23), 1.64(0.14), 0.61(0.12), –0.74(0.13), –2.15(0.17), –3.93(0.24)
		ISC9	1.67(0.14)	3.95(0.24), 1.81(0.15), 0.38(0.13), –0.74(0.13), –2.23(0.15), –3.89(0.26)
		ISC13	1.72(0.16)	2.62(0.20), 1.27(0.13), 0.05(0.12), –0.99(0.14), –2.39(0.18), –3.98(0.25)
		ISC15	1.75(0.16)	4.45(0.30), 2.28(0.19), 0.74(0.13), –0.33(0.13), –1.90(0.17), –3.88(0.26)
		ISC17	1.48(0.14)	3.03(0.21), 1.25(0.14), –0.02(0.11), –1.36(0.13), –2.90(0.19), –4.68(0.35)
		ISC19	1.16(0.14)	2.63(0.15), 1.54(0.12), 0.76(0.11), –0.32(0.10), –1.72(0.14), –2.99(0.18)
Unfavorable Comparison with Others – USC		USC2	2.22(0.21)	2.91(0.23), 0.78(0.16), –0.62(0.16), –1.76(0.18), –3.13(0.24), –4.76(0.36)
		USC4	1.12(0.13)	3.21(0.20), 1.84(0.14), 0.49(0.12), –0.17(0.12), –1.26(0.13), –2.64(0.20)
		USC10	1.72(0.17)	1.80(0.17), 0.29(0.13), –0.91(0.12), –1.70(0.14), –2.87(0.18), –3.96(0.26)
		USC14	0.52(0.12)	3.13(0.19), 1.55(0.10), 0.49(0.09), –0.25(0.09), –1.30(0.11), –2.27(0.15)
		USC18	1.52(0.15)	1.43(0.15), –0.05(0.12), –1.14(0.13), –1.89(0.15), –3.28(0.20), –4.88(0.34)
Favorable Comparison with Others – FSC		FSC8	1.23(0.13)	3.17(0.22), 1.57(0.15), 0.15(0.11), –1.06(0.13), –2.37(0.17), –3.55(0.25)
		FSC12	2.21(0.27)	2.24(0.23), 0.41(0.15), –0.68(0.17), –1.76(0.21), –2.98(0.27), –4.32(0.38)
		FSC16	1.13(0.15)	2.97(0.20), 1.40(0.12), 0.36(0.09), –0.86(0.10), –1.96(0.13), –3.55(0.24)
		FSC20	1.09(0.13)	2.68(0.17), 1.22(0.11), 0.25(0.10), –0.70(0.11), –1.72(0.14), –2.91(0.21)
		FSC21	1.74(0.17)	2.12(0.20), 0.73(0.15), –0.32(0.14), –1.59(0.16), –2.95(0.21), –4.65(0.31)

Notes. LOSC = The Levels of Self-Criticism Scale. LOSC ISC = Internalized self-criticism. LOSC USC = Unfavourable comparison with others. LOSC FSC = Favourable comparison with others.

due to its low factor loading). Test information curve (Figure 1D) for this dimension shows that it measures better the higher parts of latent ability (self-criticism). Item information curves for the dimension Favorable Comparison with Others (Figure 1E) show that items 12 and 21 have large information, and items 8, 16 and 20 have low information. Test information curve (Figure 1F) for this dimension shows that it measures better the higher parts of latent ability (self-criticism).

The bi-factor model (Reise, 2012) was fitted to test the possibility of the use of the total score. The values of the indexes of fit for the bi-factor IRT Samejima graded response model were as follows: CFI = .96, TLI = .93, RMSEA = .049 (90% confidence interval .036–.061), SRMR = .068, AIC = 38386, BIC = 39053. The likelihood-ratio test (and lower information criteria) showed that the bi-factor model had better fit than the 3-factor model, $\chi^2(16) = 164$, $p < .001$. However, the hierarchical omega is .54 which means that only 54% of the explained variance is accounted for by the single general factor. It means that the total score cannot be reliably used.

Reliability analysis

The reliability values for the Internalized Self-Criticism subscale of the original version of the LOSC (with all 22 items) and the revised version of the LOSC (without items 6, 11, and 22, which showed undesirable psychometric properties in the Slovak version) were .85 and .87, respectively. The Comparative Self-Criticism subscale had a value of .58, while Unfavourable Comparison with Others had a value of .71, and Favourable Comparison with Others had a value of .73 on the

polychoric Cronbach α test. Obviously, the values are higher after separating the factor into two factors. However, the values are just acceptable.

Validity analysis

The validity of the LOSC was estimated through Spearman's correlations between its scales and other instruments measuring related constructs (Table 4). In the case of the FSCRS, it was related to self-criticism in the form of Hated Self and Inadequate Self as similar constructs, and Reassured Self as the opposite construct. In the case of the SCS, we tested the relationship with Self-Uncompassionate responding which is the negative, opposite part of self-compassion closely reminiscent of self-criticism and Self-Compassionate responding, which should have a negative relationship to self-criticism. The correlations confirmed the theoretical expectations that all of the self-criticism factors of the LOSC would be positively correlated with any kinds of self-criticism from the FSCRS (Hated and Inadequate Self) and with self-uncompassionate responding from the SCS and its dimensions, which are Isolation, Self-Judgement, and Over-Identification. Also, the correlations confirmed the theoretical expectations that all self-criticism factors of the LOSC would be negatively correlated with self-compassionate responding and its dimensions, which are Mindfulness and Self-Kindness. Common Humanity was correlated positively with the LOSC dimension Favourable Comparison with Others. Also, the FSC and USC have opposite values in Spearman correlations, which suggests that they might be meaningfully separated into two
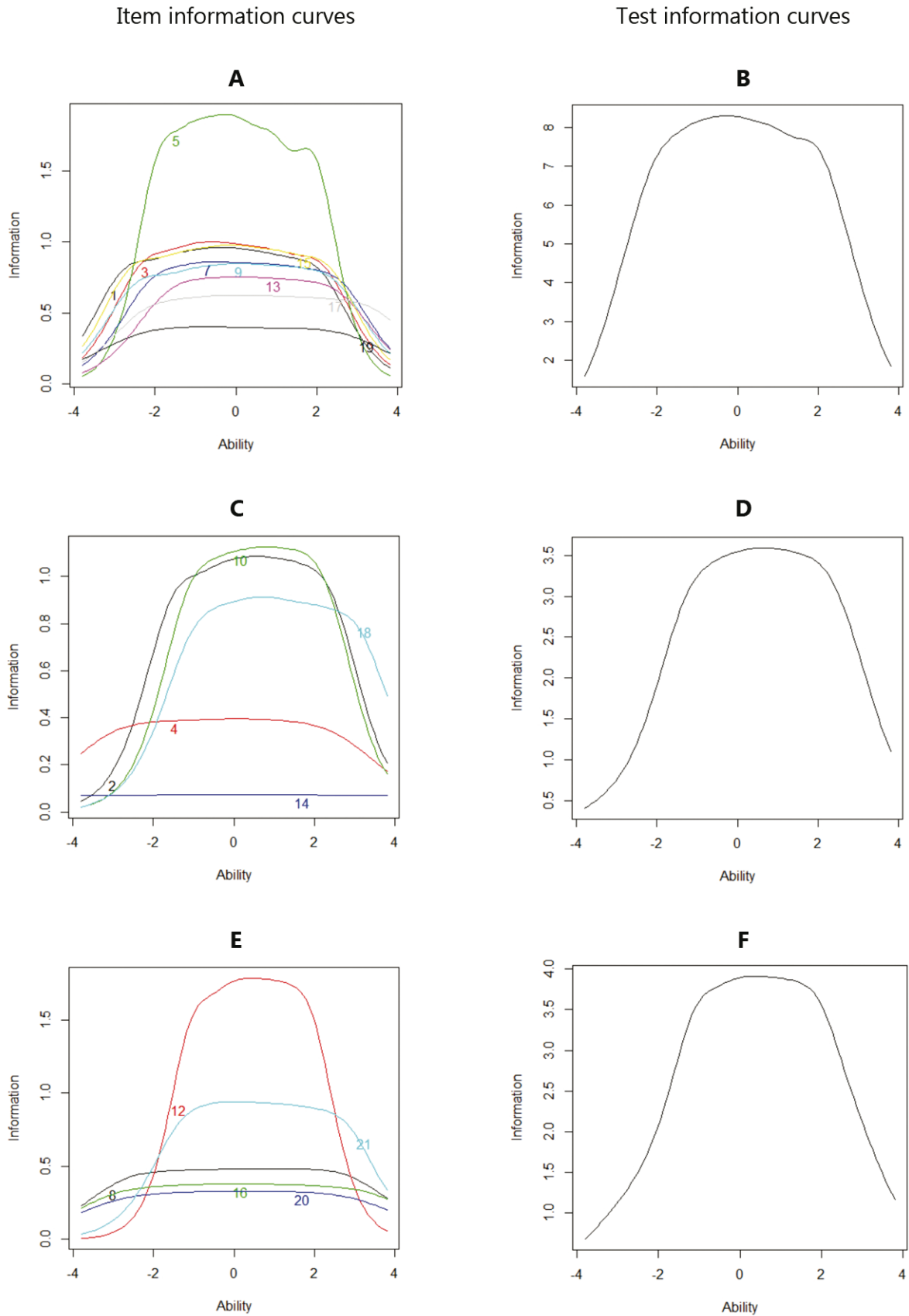
Item information curves

Test information curves

A

B

C

D

E

F

Figure 1. Item information curves for the LOSC subsales Internalized Self-Criticism (A), Unfavorable Comparison with Others (C), Favorable Comparison with Others (E), and test information curves for the LOSC subscales Internalized Self-Criticism (B), Unfavorable Comparison with Others (D), and Favorable Comparison with Others (F).

Table 4. *Nonparametric Spearman correlations between FSCRS, SCS, LOSC and its subscales*

		LOSC ISC	LOSC USC	LOSC FSC
FSCRS	Hated Self	.36***	.39***	−.31***
FSCRS	Inadequate Self	.63***	.57***	−.12**
FSCRS	Reassured Self	−.09*	−.13**	.51***
SCS	Self-compassionate responding	−.10*	−.10*	.42**
SCS	Self-uncompassionate responding	.58***	.55***	−.06
SCS	Self-kindness	−.19***	−.13***	.29**
SCS	Self-judgement	.49***	.43***	−.10*
SCS	Common humanity	.05	.01	.25***
SCS	Isolation	.48***	.51***	−.08
SCS	Mindfulness	−.10*	−.13**	.32***
SCS	Over-identification	.53***	.49***	−.06

Notes. LOSC = The Levels of Self-Criticism Scale. LOSC ISC = Internalized self-criticism. LOSC USC = Unfavourable comparison with others. LOSC FSC = Favourable comparison with others.
N = 592. * p < .05, ** p < .01, *** p < .001.

distinctive factors. However, there is still a possibility that it is just a psychometric artefact of grouping positive and negative items separately (Weijters, Baumgartner, & Schillewaert, 2013). The USC correlates positively with self-critical subscales and negatively with self-compassionate subscales, and the FSC correlates negatively with self-critical subscales and positively with self-compassionate subscales. Magnitude of correlations is generally higher for subscales with similar content (positive correlations are higher in magnitude than negative correlations). To conclude, the results indicated that the LOSC and its subscales showed promising validity.

Discussion

The goal of this paper was to translate the Levels of the Self-Criticism Scale (LOSC; Thompson & Zuroff, 2004) into Slovak and to analyse its psychometric properties and factor structure. Except for Thomson and Zuroff (2004), who created it, a factor analysis of the LOSC has never been tested. Therefore, the present paper contributes to the discussion of the elements of self-criticism and how it should be defined.

Given the size of the sample (592 respondents), we could use the item response theory (IRT) model. IRT analysis showed that the three main dimensions of the instrument (Internalized Self-Criticism, Unfavourable Comparison with Others and Favourable Comparison with Others), meet the conditions of good fit with the data, and their general factors explain a sufficient proportion of the variance, as well as the psychometric properties of the individual items which are very acceptable. However, we decided to omit items 6, 11, and 22 from the Slovak version of the scale because their descriptive analysis showed inadequate psychometric properties.

Unlike the original study by Thompson and Zuroff (2004) which used a two-dimensional model, a three-factor model proved to be more appropriate in our sample. However, the internalized factor of Self-Criticism remained the same. Only the factor of Comparative Self-Criticism on this sample was separated into two smaller factors of Unfavourable Comparison with Others and Favourable Comparison with Others. Participants in this study answered differently to the items formulated in a positive way (e.g., "I am confident that most of the people I care about will accept me for who I am.") and those described in a negative manner (e.g. "I have a nagging sense of inferiority."). The positively formulated items were grouped in the factor of Favourable Comparison with Others and the negatively formulated items were grouped in the factor of Unfavourable Comparison with Others. In fact, this finding about dividing Comparative Self-Criticism into Favourable Comparison with Others and Unfavourable Comparison with Others might be merely a psychometric artefact: there could be a single Comparative Self-Criticism after all, but the LOSC is unable to measure it reliably. This could be due to the previous findings which demonstrated that reverse-scored items very often cluster into a separate factor (Carlson et al., 2011). Thus, we have to be careful in stating that there are two separate factors as they had very high correlations between them. Given that none of the studies using the LOSC conducted IRT factor analysis, we cannot compare the results further. It is difficult to fully compare results of factors analyses based on different methods of estimation (IRT methods versus linear methods). This issue must be resolved by subsequent research. For example, it would be beneficial to re-analyse the original data set from Thomson and Zuroff (2004) using a more advanced statistical approach, i.e. item response theory, and compare the results with the Slovak results because there is no such a study yet.

In the Slovak version, we found at least an acceptable internal consistency of each revised dimension (.71, .73, .87), as measured by Cronbach's alpha. The coefficient for original first dimension (.85) was approximately the same for the first dimension as in the original research study (Thomson & Zuroff, 2004) and the previous research findings (Smart et al., 2016; Yamaguchi et al., 2014), but lower (.58) for the second dimension. However, the reliability is generally acceptable after dividing the scale into three factors.

To verify the validity of the LOSC, the previous study (Thompson & Zuroff, 2004) also used distant constructs, such as self-esteem as well as the same construct as the self-criticism subscale measured by The Depressive Experiences Questionnaire (DEQ; Blatt et al., 1976). Positive correlations of the LOSC subscales were found with the self-criticism

subscale of the DEQ, perfectionism and negative correlations with self-esteem (Thompson & Zuroff, 2004). In this study, we tested the validity of the LOSC with all of the existing scales which measure primary self-criticism or self-reassurance and respectively self-compassion. The expected relationships in the SCS and FSCRS subscales were confirmed in terms of a positive correlation with Self-Criticism and a negative correlation with Self-Compassion and Self-Reassurance. In addition, the USC correlates positively with self-critical subscales and self-uncompassionate subscales and negatively with self-compassionate subscales, and the FSC correlates negatively with self-critical subscales and self-uncompassionate subscales and positively with self-compassionate subscales, which suggests that they could be meaningfully separated into two distinctive factors. However the possibility that it is just a psychometric artefact of grouping positive and negative items of the LOSC separately (Weijters et al., 2013) still remains open.

The study is mainly limited by the fact that our sample is not representative of the Slovak population. Although the psychometric findings about the Slovak version of the LOSC are generally acceptable, it is imperative to explore the LOSC in a larger and more representative sample in Slovakia. Further attention will require different age and education groups as well as diverse regions of the country and personality traits to see if these variables interact with the level of self-criticism.

Self-criticism is considered to be the general underlying factor of various kinds of psychopathology (Falconer et al. , 2015). Therefore, in future research, a sample should be extended to a clinical population with various diagnoses in order to create norms and upgrade diagnostics by distinguishing individuals suffering from pathological self-criticism from the individuals with high self-criticism within the non-clinical population. Future studies could entail an expert assessment of the self-criticism level or measure the physiological indicators of self-criticism such as heart rate variability (Rockliff, Gilbert, McEwan, Lightman, & Glover, 2008) to provide additional evidence of validity.

Conclusion

Despite the limitations of the non-representative sample, the Slovak version of the LOSC appears to have an acceptable level of reliability and validity after omitting three items. Therefore, the scale has the potential to be used in the Slovak context to assess the level of self-criticism in the community population. Compared to the English version with two dimensions (Internalized Self-Criticism and Comparative Self-Criticism) using classical factor analysis, the Slovak translation has three factors (Internalized Self-Criticism, Unfavourable Comparison with Others and Favourable Comparison with Others) exploiting more advanced IRT factor analysis. As there is no theoretical explanation for the three factors solution, it might be merely a result of a psychometric artefact because positive and negative items have tendency to split into different factors (Weijters et al., 2013). In the future, the use of the scale in a clinical setting and the factor analysis of the scale in different language translations will have to be verified.

References

Anscombe, F. J., & Glynn, W. J. (1983). Distribution of kurtosis statistic for normal statistics. *Biometrika, 70*(1), 227–234.

Bartholomew, K., & Horowitz, L. M. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality and Social Psychology, 61*, 226–244.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin, 112*, 400–404.

Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. *International Journal of Exercise Science, 8*(3), 297–302.

Blatt, S. J. (1974). Levels of object representation in anaclitic and introjective depression. *Psychoanalytic Study of the Child, 24*, 107–157.

Blatt, S. J., D'Afflitti, J. P., & Quinlan, D. M. (1976). *Depressive Experiences Questionnaire*. Unpublished manuscript, Yale University, New Haven, CT, USA.

Bowlby, J. (1980). *Attachment and loss: Vol. 3: Loss, separation, and depression*. New York, NY, USA: Basic Books.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA, USA: Sage.

Cai, L. (2010). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics, 35*, 307–335.

Carlson, M., Wilcox, R., Chou, C. P., Chang, M., Yang, F., Blanchard, J., Marterella, A., Kuo, A., & Clark, F. (2011). Psychometric properties of reverse-scored items on the CES-D in a sample of ethnically diverse older adults. *Psychological Assessment, 23*, 558–562.

Castilho, P., Pinto-Gouveia, J., & Duarte J. (2015). Exploring Self-criticism: Confirmatory Factor Analysis of the FSCRS in Clinical and Nonclinical Samples. *Clinical Psychology & Psychotherapy, 22*(2), 153–164.

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48*(6), 1–29.

Chen, S. X., Lam, B., Wu, W., Ng, J., Buchtel, E., Guan, Y., & Deng, H. (2016). 'Do people's world views matter? The why and how.' *Journal of Personality and Social Psychology, 110*(5), 743–765.

Costa, P. T. & McCrae, R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Model (NEO-FFI)*. Professional manual. Odesa, FL, USA: Psychological Assesment Center.

Crăciun, B. (2013). Coping strategies, self-criticism and gender factor in relation to quality of life. *Procedia - Social and Behavioral Sciences, 78*, 466–470.

Cucina, J. & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence, 5*(3), 27–41.

D'Agostino, R. B. (1970). Transformation to normality of the null distribution of G1. *Biometrika, 57*(3), 679–681.

Danielsson, G. (2012). *En psykometrisk utprövning av The Levels of Self-Criticism Scale (LOSC) [A psychometric test of The Levels of Self-Criticism Scale (LOSC)]*. Lund, Norway: Psykoterapeutexamensuppsats, Lunds universitet, Institutionen för psykologi.

Dinno, A. (2009). Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. *Multivariate Behavioral Research, 44*, 362–388.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412.

Falconer, C. J., King, J.A., & Brewin, C. R. (2015). Demonstrating mood repair with a situation-based measure of self-compassion and self-criticism. *Psychology and Psychotherapy, 88*(4), 351–365.

Gel, Y. R., & Gastwirth, J. L. (2008). A robust modification of the Jarque-Bera test of normality. *Economics Letters, 99*, 30–32.

Ghorbani, N., Mousavi, A., Watson, P. J., & Chen, Z. (2011). Integrative self-knowledge and the harmony of purpose model in Iranian autoimmune patients. *Electronic Journal of Applied Psychology, 7*(2), 1–8.

Gilbert, P., & Choden. (2013). *Mindful compassion*. London, United Kingdom: Constable-Robinson.

Gilbert, P., Clark, M., Hempel, S., Miles, J. N. V., & Irons, C. (2004) Criticising and reassuring oneself: An exploration of forms, styles and reasons in female students. *British Journal of Clinical Psychology, 43*, 31–50.

Gilbert, P., & Irons, C. (2009). Shame, self-criticism, and self-compassion in adolescence. In N. B. Allen (Ed.), *Adolescent emotional development and the emergence of depressive disorders* (pp. 195–214). Cambridge, United Kingdom: Cambridge University Press.

Greenberg, L. S. (2011). *Emotion-focused therapy*. Washington DC, USA: American Psychological Association.

Halamová, J., & Kanovský, M. (2017). *Sebasúcit a sebakritickosť: Psychometrická analýza meracích nástrojov [Self-criticism and self-blame: Psychometric analysis of measuring instruments]*. Bratislava, Slovakia: Univerzita Komenského v Bratislave.

Halamová, J., Kanovský, M., & Pacúchová, M. (2017). Robust psychometric analysis and factor structure of The Forms of Self–criticizing/Attacking and Self–reassuring Scale. *Československá psychologie, 61*(4), 331–349.

Halamová, J., Kanovský, M., and Pacúchová, M. (2018). Self-Compassion Scale: IRT Psychometric Analysis, Validation, and Factor Structure – Slovak Translation. *Psychologica Belgica, 57*(4), 190–209.

Hewitt, P. L. & Flett, G. (1989). The Multidimensional Perfectionism Scale: development and validation. *Canadian Psychology, 30*, 339.

Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.

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.

Levinger, G., & Pietromonaco, P. (1989). *Conflict Style Inventory*. Unpublished scale, University of Massachusetts, Amherst.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949.

Manfredi, C., Caselli, G., Pescini, F., Rossi, M., Rebecchi, D., Ruggiero, G. M., & Sassaroli S. (2016). Parental criticism, self-criticism and their relation to depressive mood: An exploratory study among a non-clinical population. *Research in Psychotherapy: Psychopathology, Process and Outcome, 19*(1), 41–48.

Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). Comparing the fit of item response theory and factor analysis models. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(3), 333–356.

McDonald, R. P. (1997). Normal-ogive multidimensional model. In W.J. Linden, & Hambleton R. K. (Eds.), *Handbook of modern item response theory* (pp. 258–270). New York, NY, USA: Springer.

Neff, K. D. (2003a). The development and validation of a scale to measure self-compassion. *Self and identity, 2*(3), 223–250.

Neff, K. D. (2003b). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity, 2*, 85–101.

Neff, K. D., & Dahm, K. A. (2015). Self-compassion: What it is, what it does, and how it relates to mindfulness. In M. Robinson, B. Meier, & B. Ostafin (Eds.) *Mindfulness and self-regulation* (pp. 1–40). New York, NY, USA: Springer.

Öngen, D. (2006). The relationships between self-criticism, submissive behavior and depression among Turkish adolescents. *Personality and Individual Differences, 41*(5), 793–800.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from: http://www.R-project.org/

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*, 667–696.

Revelle, W. (2015). *psych: Procedures for Personality and Psychological Research*. Retrieved from: http://CRAN.R-project.org/package=psych

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373.

Rockliff, H., Gilbert, P., McEwan, K., Lightman, S., & Glover, D. (2008). A pilot exploration of heart rate variability and salivary cortisol responses to compassion-focused imagery. *Clinical Neuropsychiatry: Journal of Treatment Evaluation, 5*(3), 132–139.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ, USA: Princeton University Press.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34*(4, Pt. 2), 100.

Samejima F. (1974). Normal ogive model for the continuous response level in the multidimensional latent space. *Psychometrika*, *39*, 111–121.

Smart, L. M., Peters, J. R., & Baer, R. A. (2016). Development and validation of a measure of self-critical rumination. *Assessment*, *23*(3), 321–332.

Thompson, R., & Zuroff, C. (2004). The Levels of Self-Criticism Scale: Comparative self-criticism and internalized self-criticism. *Personality and Individual Differences*, *36*(2), 419–430.

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, *18*(3), 320–334.

Yamaguchi, A., & Kim, M. S. (2013). Effects of self-criticism and its relationship with depression across cultures. *International Journal of Psychological Studies*, *5*, 1.

Yamaguchi, A., Kim, M. S., & Akutsu, S. (2014). The effects of self-construals, self-criticism, and self-compassion on depressive symptoms. *Personality and Individual Differences*, *68*, 65–70.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients Alpha and Theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*, 21–29.