Original Article

Longitudinal Measurement Invariance of the Brief Symptom Inventory (BSI)-18 in Psychotherapy Patients

Ruth von Brachel,[1] Angela Bieda,[2] Jürgen Margraf,[1] and Gerrit Hirschfeld[2]

[1]Mental Health Research & Treatment Center of Ruhr-University Bochum, Germany
[2]Faculty of Business Management and Social Sciences, University of Applied Sciences Osnabrück, Germany

Abstract: The Brief Symptom Inventory (BSI)-18 is a widely-used tool to assess changes in general distress in patients despite an ongoing debate about its factorial structure and lack of evidence for longitudinal measurement invariance (LMI). We investigated BSI-18 scores from 1,081 patients from an outpatient clinic collected after the 2nd, 6th, 10th, 18th, and 26th therapy session. Confirmatory factor analysis (CFA) was used to compare models comprising one, three, and four latent dimensions that were proposed in the literature. LMI was investigated using a series of model comparisons, based on chi-square tests, effect sizes, and changes in comparative fit index (CFI). Psychological distress diminished over the course of therapy. A four-factor structure (depression, somatic symptoms, generalized anxiety, and panic) showed the best fit to the data at all measurement occasions. The series of model comparisons showed that constraining parameters to be equal across time resulted in very small decreases in model fit that did not exceed the cutoff for the assumption of measurement invariance. Our results show that the BSI-18 is best conceptualized as a four-dimensional tool that exhibits strict longitudinal measurement invariance. Clinicians and applied researchers do not have to be concerned about the interpretation of mean differences over time.

Keywords: Brief Symptom Inventory (BSI)-18, confirmatory factor analysis, longitudinal measurement invariance, psychological distress

The main goal of psychotherapy and psychiatry is to reduce psychological distress and to improve mental health. Physicians use measures of psychological distress to identify patients in need for psychological interventions, therapists measure psychological distress during treatment to monitor their patients' progress, and researchers measure the development of psychological symptoms over time as a crucial aspect of any clinical trial on psychological interventions. In addition to scales tailored to specific conditions or symptoms, several broad symptom scales have been developed to assess a broader number of symptom domains. The Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983) and its short forms (Prinz et al., 2013) measure overall distress of a person as well as individual symptoms such as anxiety, depression, and somatization. The BSI may be used to identify those medical patients who also need psychological help (Zabora et al., 2001), to monitor the progress during therapy (Geisheim et al., 2002), or to measure outcomes in clinical studies (Piersma, Reaume, & Boes, 1994). Since monitoring progress and measuring outcomes involve comparisons across time, it is vital to establish the longitudinal measurement invariance (LMI). The aims of the present study were twofold, first to investigate the factorial structure of the BSI in a large sample of outpatients and second to test for its LMI.

There is an ongoing debate about the underlying structure of the BSI and its short form the BSI-18 (for reviews see, Loutsiou-Ladd, Panayiotou, & Kokkinos, 2008; Prinz et al., 2013). The BSI-18 was developed to assess three central symptom domains (somatization, depression, and anxiety) with six items each (Derogatis, 2000). Several studies conclude that these three scales have good psychometric properties and responses conform to this factorial structure (Abraham, Gruber-Baldini, Harrington, & Shulman, 2017; Derogatis, 2000; Durá et al., 2006; Galdón et al., 2008; Petkus et al., 2010; Prinz et al., 2013; Recklitis et al., 2006; Spitzer et al., 2011; Torres, Miller, & Moore, 2013; Wang et al., 2010; Wiesner et al., 2010; Zabora et al., 2001). There are, however, also authors suggesting that this measure assesses a single "psychological distress" factor (Asner-Self,

Schreiber, & Marotta, 2006; Prelow, Weaver, Swenson, & Bowman, 2005), and other authors claiming that the BSI-18 is best conceptualized as measuring four factors (Andreu et al., 2008; Zabora et al., 2001). The four-factor models retain the somatization and depression factors but split the anxiety scale into two factors, that is, generalized distress and panic, each assessed by three items. Resolving this issue has important consequences for the interpretation of scale scores and is an important prerequisite for testing LMI.

Our second aim is to test the LMI of the BSI-18. LMI is a special case of measurement invariance testing. While measurement invariance tries to answer the question whether a specific scale has the same meaning to different groups of participants, LMI tries to answer the question whether a scale has the same meaning over time (Widaman, Ferrer, & Conger, 2010). Even though LMI seems to be a fundamental aspect of psychological scales it is rarely tested (Borsboom, 2006). Several recent studies found that frequently used depressions scales such as the Beck Depression Inventory and the Hamilton Rating Scale for Depression lack LMI (Fokkema, Smits, Kelderman, & Cuijpers, 2013; Fried et al., 2016; Wu, 2016) while the Center for Epidemiologic Studies Depression Scale exhibits LMI (Ferro & Speechley, 2013). Thus, LMI has to be established for each individual scale in sufficiently large samples.

Materials and Methods

Participants

Participants in this study were outpatients at the Mental Health Research and Treatment Center at the psychology department of the Ruhr-University Bochum between the years 1990 and 2012 at which time the BSI was removed as a standard assessment. All therapists at this facility had at least a Master degree in psychology and at least one year of practical training in cognitive-behavioral therapy (CBT). All participants were undergoing CBT for a variety of diagnoses. Participants were asked to fill out questionnaires assessing their general mental health, their symptoms, and their satisfaction with their therapy at different points in treatment as part of routine diagnostic sessions. These sessions were scheduled after the 2nd, 6th, 10th, 18th, and 26th session. This study was approved by the local ethics committee under the number 318.

The Brief Symptom Inventory-18

The BSI (Derogatis & Melisaratos, 1983) in its original format measures symptom severity in nine different domains (Somatization, Obsession-Compulsion, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Paranoid Ideation, Phobic Anxiety, and Psychoticism). There is a variety of short forms with the 18-item version (Derogatis, 2000) showing the best psychometric properties (Prinz et al., 2013). The BSI-18 consists of a 5-point Likert scale, participants indicate their agreement from 0 = *not at all* to 4 = *extremely*.

The psychometric properties of the German version of the BSI-18 are good with internal consistencies ranging from .63 to .93 and item-total correlation $\geq$ 0.40. It is also correlated moderately to highly with other measures of symptom severity (Prinz et al., 2013; Spitzer et al., 2011).

Data Analysis

Data were analyzed in three steps. First, we inspected individual items over time. Second, we used confirmatory factor analysis (CFA) to decide on the model that should be tested for measurement invariance. Specifically, we tested – for the five measurement occasions separately – the three different models that are discussed in the literature. In all models, each item loaded on only one latent variable and latent variables were correlated with one another. In keeping with studies into the factor structure of the BSI-18 (Recklitis et al., 2006; Torres et al., 2013; Wang et al., 2010; Wiesner et al., 2010) model parameters were estimated using Maximum Likelihood (ML) with robust standard errors to account for non-normality. We used conventional fit indices to assess overall model fit (Hu & Bentler, 1999), Akaike Information Criterion (AIC) to compare models at the individual time-points and report standardized loadings and covariances.

Third, we systematically assessed LMI by fitting a series of more and more restricted models to the data. The baseline (configural) model for this analysis was based on the parameterization developed by Widaman and colleagues (2010) for testing for LMI. In this model 20 latent variables (four at each of the five measurement occasions) were used to model the responses to the 90 (18 items at 5 measurement occasions) observed variables. In the first step this parameterization entails first setting the mean of the latent variables at the first measurement occasion to 0 and the variance to 1. Second, the loadings of the first items on each latent variable were freely estimated and corresponding loadings of the first items were constrained to be equal across time. Third, the intercepts of the first items on each latent variable were estimated freely but the corresponding intercepts were constrained to be equal across time. The other parameters, that is, intercepts and variances for the latent variables, covariances among latent variables assessed at one measurement occasion, and covariances between items at different measurement occasions, were

estimated freely. The weak invariance model added across-time invariance constraints on the remaining loadings. Since 4 of the 18 loadings were already constrained in the baseline model this yielded 56 ($18 - 4 \times 4$) degrees of freedom. The strong invariance model added across-time invariance constraints on the item intercepts. Again, since 4 of the 18 item intercepts were already constrained, this yielded 56 degrees of freedom. Lastly, the strict invariance model further added constraints on the residual variances of the items yielding 56 degrees of freedom. These different models were first compared using scaled chi-square tests (Satorra & Bentler, 2001). We calculate the effect size w that is based on chi-square test to describe the magnitude of the invariance. Because w is equal to Pearsons' correlation coefficient it may be interpreted using the same conventions for small (w = 0.1), medium (w = .3), and large (w = .05) effect sizes (Newsom, 2015, p. 30). Since these chi-square based statistics are sensitive to sample size (Cheung & Rensvold, 2002; Little, 1997), we also used changes in fit indices to describe model differences. Specifically, we interpreted increases in CFI < .01 as indicating a similarly good model fit and thus support for invariance assumption (Cheung & Rensvold, 2002). All analyses were performed in R using the lavaan package (Rosseel, 2012).

Results

Sample

Overall, 1,771 patients visited the clinic at least once and participated in diagnostic sessions during the study period. Of these 590 patients took part in less than five diagnostic sessions, indicating short-term therapies or treatment termination. Of the 1,181 patients with at least five BSI-scores, no demographic or clinical data could be found for 100, leaving 1,081 patients that were included in the analysis. There were no statistically significant differences in the overall BSI-score at baseline between included and excluded participants. Participants were on average 37.07 years old (min = 18; max = 69; SD = 10.64) and mostly female (n = 632; 58%). The five most frequent first diagnoses according to *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)* terminology were panic disorder (n = 187; 17%), social phobia (n = 168; 15%), major depression with recurrent (n = 128; 12%) or a single episode (n = 76; 7%), and obsessive-compulsive disorder (n = 57; 5%).

Descriptive Statistics

As a first step, we calculated the individual item means for each measurement occasion. As can be seen in Table 1 participants showed decreased ratings in later compared to earlier sessions.

Structure of the BSI-18

To determine the structure of the BSI-18 and to yield a valid baseline model for invariance testing, we fitted the one-, three- and four-dimensional models described above to each of the five measurement occasions separately. The fit indices were good for all models and showed that the best fitting model according to AIC was always the model with four factors. Model fit for the four-factor model was good for all measurement occasions (Table 2). All individual standardized loadings were substantial (> .30) and increased for most items from the first to the last measurement occasion (see Table S1 in the Electronic Supplementary Material, ESM 1). Correlations between factors were substantial in both the three ($.48 < r < .79$) and four-factor model ($.48 < r < .83$; see Tables S2 and S3 in ESM 1). This four-factor solution was also invariant across gender and age (see S4 in ESM 1).

Longitudinal Measurement Invariance

The series of model comparisons showed that adding constraints to the model yielded significantly decreased model fit as measured by chi-square tests (see Table 3), that is, constraining the loadings to be equal across time resulted in a significant decrease in model fit, $\chi^2(56) = 455.8$; $p < .001$; w = 0.087; ΔCFI = 0.0041. However, these decreases were all smaller than the threshold of ΔCFI < .01. Similarly, the comparisons for strong and strict invariance showed very small albeit significant differences between the constrained and unconstrained models.

Discussion

In the present study, the factorial structure and LMI of the BSI-18 during psychotherapy was investigated in a large outpatient sample. We found that a model with four latent dimensions showed the best fit to the data for all five measurement occasions. The latent dimensions were depression, somatization and two anxiety dimensions relating to panic and general anxiety. We also found that the BSI-18 exhibits strict longitudinal invariance. In the following, we will discuss each of these findings in turn, before describing general methodological aspects and limitations of the present study.

An open issue concerning the factor structure of the BSI-18 is whether or not the anxiety factor needs to be split up into two factors (general anxiety and panic).

Table 1. Item means for all measurement occasions

Item	T1	T2	T3	T4	T5
Nervousness	2.90 (1.20)	2.71 (1.12)	2.55 (1.10)	2.47 (1.08)	2.41 (1.08)
Scared	1.92 (1.15)	1.77 (1.04)	1.68 (0.94)	1.63 (0.92)	1.64 (0.92)
Lonely	2.85 (1.37)	2.50 (1.28)	2.41 (1.24)	2.31 (1.24)	2.28 (1.21)
Blue	2.75 (1.25)	2.52 (1.17)	2.42 (1.17)	2.36 (1.21)	2.29 (1.15)
No interest	2.53 (1.34)	2.27 (1.23)	2.18 (1.23)	2.07 (1.20)	1.99 (1.13)
Fearful	2.44 (1.27)	2.24 (1.18)	2.13 (1.14)	2.05 (1.12)	2.00 (1.06)
Faintness	1.90 (1.09)	1.79 (1.00)	1.74 (0.96)	1.70 (0.96)	1.69 (0.93)
Nausea	2.08 (1.20)	2.03 (1.18)	1.98 (1.14)	1.91 (1.09)	1.87 (1.08)
Short of breath	1.89 (1.14)	1.77 (1.03)	1.74 (1.05)	1.69 (0.96)	1.68 (0.99)
Numb/tingling	1.76 (1.09)	1.66 (0.98)	1.69 (1.00)	1.66 (0.98)	1.66 (0.97)
Hopelessness	2.96 (1.39)	2.67 (1.32)	2.51 (1.27)	2.43 (1.27)	2.37 (1.26)
Body weakness	2.11 (1.20)	2.02 (1.17)	1.93 (1.14)	1.87 (1.10)	1.86 (1.09)
Tense	3.09 (1.25)	2.91 (1.18)	2.81 (1.15)	2.72 (1.13)	2.66 (1.15)
Panic episodes	2.32 (1.42)	1.97 (1.19)	1.89 (1.12)	1.76 (0.99)	1.78 (1.03)
Restlessness	2.23 (1.26)	2.13 (1.17)	2.11 (1.17)	2.02 (1.14)	1.97 (1.12)
Worthlessness	2.78 (1.45)	2.49 (1.36)	2.33 (1.33)	2.25 (1.31)	2.22 (1.28)
Chest pains	1.84 (1.10)	1.76 (1.01)	1.72 (0.99)	1.67 (0.92)	1.69 (0.97)
Suicidal thoughts	1.73 (1.10)	1.53 (0.92)	1.50 (0.88)	1.49 (0.86)	1.52 (0.91)

Table 2. Model Fit Indices at the five measurement occasions

Time	Factors	χ^2	*df*	CFI	AIC	RMSEA	SRMR
1	One	1,842.351	135	0.724	58,051.912	0.120	0.089
1	Three	551.048	132	0.933	56,475.422	0.060	0.045
1	Four	482.162	129	0.943	56,397.638	0.055	0.043
2	One	1,680.569	135	0.766	53,671.268	0.119	0.079
2	Three	572.732	132	0.965	35,303.823	0.055	0.037
2	Four	456.591	129	0.951	52,038.380	0.055	0.038
3	One	1,522.678	135	0.788	52,296.144	0.115	0.077
3	Three	509.344	132	0.944	50,872.420	0.060	0.043
3	Four	435.645	129	0.955	50,778.330	0.054	0.041
4	One	1,518.242	135	0.803	50,542.507	0.114	0.075
4	Three	507.702	132	0.948	49,143.787	0.060	0.042
4	Four	417.124	129	0.960	49,024.902	0.053	0.040
5	One	1,383.608	135	0.807	49,982.899	0.113	0.074
50	Three	451.223	132	0.952	48,571.511	0.057	0.037
5	Four	383.586	129	0.962	48,478.828	0.051	0.036

Notes. *df* = degrees of freedom; CFI = Comparative Fit Index; AIC = Akaike Information Criterion; RMSEA = Root Mean Squared Error of Approximation; SRMR = Standardized Root Mean squared Residual.

Unfortunately, the authors that have contributed to this literature not only studied different populations but used vastly different methods to decide for or against a specific number of factors such as parallel analysis, EFA, and CFA. Studies supporting a three-dimensional structure did not explicitly test for the number of dimensions to retain, or compare the three- to the four-dimensional structure (Prinz et al., 2013; Spitzer et al., 2011; Torres et al., 2013). Studies supporting the single-factor model also did not directly compare the different models (Prelow et al., 2005) or compared a single-factor model with post hoc modifications to an unaltered three-factor model (Asner-Self et al., 2006). In the present study, we directly compared the proposed models using CFA and found that a four-factor model with two anxiety dimension fit the data at all measurement occasions better than three- or one-dimensional models. While the correlations between the panic and generalized anxiety scores were high, they were

Table 3. Model comparisons for LMI testing

Model	χ^2	*df*	RMSEA	CFI	SRMR	$\Delta\chi^2$	*w*	ΔCFI
Configural	8,593.846	3,669	0.035	0.931	0.218			
Weak	8,942.657	3,725	0.036	0.927	0.237	455.80	0.087	0.0041
Strong	9,114.393	3,781	0.036	0.926	0.240	242.25	0.064	0.0016
Strict	9,852.268	3,837	0.038	0.916	0.239	713.47	0.109	0.0095

Note. $\Delta\chi^2$ were calculated based on the scaled test statistic (Satorra & Bentler, 2001).

still smaller than suggested thresholds for factor correlations (Kline, 2015). Importantly, all other studies directly comparing the three and four-factor model found that the four-factor solution had a better fit to the data than the three-factor solution (Abraham et al., 2017; Durá et al., 2006; Galdón et al., 2008; Petkus et al., 2010; Recklitis et al., 2006; Wang et al., 2010; Wiesner et al., 2010). The authors all indicate that these small improvements may be due to chance and name theoretical reasons for their preference of the three-dimensional structure. While we believe it is very prudent to retain a factor model even if results of a single study point in a different direction, especially if it is a small study (Durá et al., 2006; Galdón et al., 2008; Hirschfeld, von Brachel, & Thielsch, 2014; Petkus et al., 2010), we believe that there is now converging evidence from several large-scale studies with several thousand participants across several countries and with diverse background (Abraham et al., 2017; Recklitis et al., 2006; Wiesner et al., 2010) all indicating that the four-dimensional model has a better fit to the empirical data than the three-factor model. This is also in-line with recent work in clinical psychology showing that panic disorder and generalized anxiety disorder are clearly distinct regarding their general distress, their psychophysiological patterns (McTeague & Lang, 2012) as well concerning their long-term course (Bruce et al., 2005). Both, however, share common etiological features such as uncertainty intolerance (Carleton et al., 2014) and are often comorbid, which may explain the high intercorrelation between the two factors in this study.

Our investigation of the LMI showed some significant decreases in model fit when LMI was assumed, the magnitude of these decreases was smaller than established thresholds (Cheung & Rensvold, 2002) and appear to be smaller than studies into the BDI (Fokkema et al., 2013). Thus, it is appropriate to use BSI-18 scores for pre-post comparisons (Fokkema et al., 2013; Newsom, 2015; Widaman et al., 2010). Since measurement invariance is a property of the measurement scale rather than the construct under investigation, it is of little use to speculate on differences to studies into LMI that used different measures.

However, reviewing other LMI studies raises a more general question about the methods that are used to investigate LMI. Researchers studying LMI (Ferro & Speechley, 2013; Fokkema et al., 2013) – and also MI (Raghavan, Rosenfeld, & Rasmussen, 2015; Torres et al., 2013; Wang et al., 2010; Wiesner et al., 2010) – differ in their choice of parameterization of the model (via loadings of items or setting latent variable variances), estimation procedure (ML, robust ML, weighted least squares), and criterion for LMI (chi-square tests, changes in CFI). While all these have an impact on the conclusions that might be drawn there are no simulation studies in the context of LMI testing to guide these decisions. In addition to establishing evidence-based guidelines for standards to test for LMI it may be instructive to take a more graded approach to invariance testing. Rather than assuming that LMI is either present – if a decrease model fit is larger than some threshold – or absent – if a decrease model fit is smaller than some threshold – it may be more useful to think about the magnitude of these differences and their effect on clinical research questions.

The present study has some limitations that need to be kept in mind. First, there is no information concerning the detailed content of the treatment (other than all received CBT), medication or duration of illness which may have influenced patients' view of their illness and consequently their answers to the BSI-18. Second, since the data has been collected over 20 years, it is also possible that cohort effect or different Zeitgeist causes differences in answering the BSI-items. Third, the sample consisted of psychotherapy outpatients in a naturalistic setting, which allows for generalization for people with mild to moderate mental illness rather than psychiatric inpatients or somatic medical conditions like cancer. It would be interesting to investigate whether the BSI-18 possesses LMI for patients in somatic medical treatments. Furthermore, replication of LMI in other cultural settings would be beneficial.

In summary, we find that the BSI-18 is best conceptualized as a four-dimensional screening instrument assessing depression, general anxiety, panic, and somatic symptoms. Furthermore, our finding of strict invariance supports the use of simple sum scores to assess individual change of symptom and distress in outpatient samples during psychotherapy.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1015-5759/a000480

ESM 1. Tables (.pdf)
Tables gives the standardized factor loadings over time (S1) as well as the factor intercorrelations for the three- (S2) and four- (S3) factor solutions and results of cross-sectional invariance tests.

References

Abraham, D. S., Gruber-Baldini, A. L., Harrington, D., & Shulman, L. M. (2017). The factor structure of the Brief Symptom Inventory-18 (BSI-18) in Parkinson disease patients. *Journal of Psychosomatic Research, 96*, 21–26. https://doi.org/10.1016/j.jpsychores.2017.03.002

Andreu, Y., Galdón, M. J., Dura, E., Ferrando, M., Murgui, S., García, A., & Ibáñez, E. (2008). Psychometric properties of the Brief Symptoms Inventory-18 (BSI-18) in a Spanish sample of outpatients with psychiatric disorders. *Psicothema, 20*, 844–850.

Asner-Self, K. K., Schreiber, J. B., & Marotta, S. A. (2006). A cross-cultural analysis of the Brief Symptom Inventory-18. *Cultural Diversity & Ethnic Minority Psychology, 12*, 367–375. https://doi.org/10.1037/1099-9809.12.2.367

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care, 44*, S176–S181.

Bruce, S. E., Yonkers, K. A., Otto, M. W., Eisen, J. L., Weisberg, R. B., Pagano, M., … Keller, M. B. (2005). Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: A 12-year prospective study. *The American Journal of Psychiatry, 162*, 1179–1187.

Carleton, R. N., Duranceau, S., Freeston, M. H., Boelen, P. A., McCabe, R. E., & Antony, M. M. (2014). "But it might be a heart attack": Intolerance of uncertainty and panic disorder symptoms. *Journal of Anxiety Disorders, 28*, 463–470.

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

Derogatis, L. R. (2000). *The Brief Symptom Inventory-18 (BSI-18): Administration, scoring, and procedures manual*. Minneapolis, MN: National Computer Systems.

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*, 595–605.

Durá, E., Andreu, Y., Galdón, M. J., Ferrando, M., Murgui, S., Poveda, R., & Jimenez, Y. (2006). Psychological assessment of patients with temporomandibular disorders: Confirmatory analysis of the dimensional structure of the Brief Symptoms Inventory 18. *Journal of Psychosomatic Research, 60*, 365–370.

Ferro, M. A., & Speechley, K. N. (2013). Factor structure and longitudinal invariance of the Center for Epidemiological Studies Depression Scale (CES-D) in adult women: Application in a population-based sample of mothers of children with epilepsy. *Archives of Women's Mental Health, 16*, 159–166.

Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response shifts in mental health interventions: An illustration of longitudinal measurement invariance. *Psychological Assessment, 25*, 520–531. https://doi.org/10.1037/a0031669

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time… or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment, 28*, 1354–1367. https://doi.org/10.1037/pas0000275

Galdón, M. J., Durá, E., Andreu, Y., Ferrando, M., Murgui, S., Pérez, S., & Ibañez, E. (2008). Psychometric properties of the Brief Symptom Inventory-18 in a Spanish breast cancer sample. *Journal of Psychosomatic Research, 65*, 533–539.

Geisheim, C., Hahlweg, K., Fiegenbaum, W., Frank, M., Schröder, B., & von Witzleben, I. (2002). Das Brief Symptom Inventory (BSI) als Instrument zur Qualitätssicherung in der Psychotherapie [The German version of the Brief Symptom Inventory (BSI): Reliability and validity in a sample of outpatient psychotherapy patients]. *Diagnostica, 48*, 28–36.

Hirschfeld, G., von Brachel, R., & Thielsch, M. (2014). Selecting items for Big Five questionnaires: At what sample size do factor loadings stabilize? *Journal of Research in Personality, 53*, 54–63. https://doi.org/10.1016/j.jrp.2014.08.003

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.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76.

Loutsiou-Ladd, A., Panayiotou, G., & Kokkinos, C. M. (2008). A review of the factorial structure of the Brief Symptom Inventory (BSI): Greek evidence. *International Journal of Testing, 8*, 90–110.

McTeague, L. M., & Lang, P. J. (2012). The anxiety spectrum and the reflex physiology of defense: From circumscribed fear to broad distress. *Depression and Anxiety, 29*, 264–281.

Newsom, J. T. (2015). *Longitudinal structural equation modeling*. New York, NY: Routledge.

Petkus, A. J., Gum, A. M., Small, B., Malcarne, V. L., Stein, M. B., & Wetherell, J. L. (2010). Evaluation of the factor structure and psychometric properties of the Brief Symptom Inventory-18 with homebound older adults. *International Journal of Geriatric Psychiatry, 25*, 578–587.

Piersma, H. L., Reaume, W. M., & Boes, J. L. (1994). The Brief Symptom Inventory (BSI) as an outcome measure for adult psychiatric inpatients. *Journal of Clinical Psychology, 50*, 555–563.

Prelow, H. M., Weaver, S. R., Swenson, R. R., & Bowman, M. A. (2005). A preliminary investigation of the validity and reliability of the Brief-Symptom Inventory-18 in economically disadvantaged Latina American mothers. *Journal of Community Psychology, 33*, 139–155.

Prinz, U., Nutzinger, D. O., Schulz, H., Petermann, F., Braukhaus, C., & Andreas, S. (2013). Comparative psychometric analyses of the SCL-90-R and its short versions in patients with affective disorders. *BMC Psychiatry, 13*, 1.

Raghavan, S. S., Rosenfeld, B., & Rasmussen, A. (2015). Measurement invariance of the Brief Symptom Inventory in survivors of torture and trauma. *Journal of Interpersonal Violence, 32*, 1708–1729. https://doi.org/10.1177/0886260515619750

Recklitis, C. J., Parsons, S. K., Shih, M.-C., Mertens, A., Robison, L. L., & Zeltzer, L. (2006). Factor structure of the Brief Symptom Inventory-18 in adult survivors of childhood cancer: results from the childhood cancer survivor study. *Psychological Assessment, 18*, 22–32. https://doi.org/10.1037/1040-3590.18.1.22

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514.

Spitzer, C., Hammer, S., Löwe, B., Grabe, H. J., Barnow, S., Rose, M., … Franke, G. H. (2011). The short version of the Brief Symptom Inventory (BSI-18): preliminary psychometric properties of the German translation. *Fortschritte der Neurologie – Psychiatrie, 79*, 517–523. https://doi.org/10.1055/s-0031-1281602

Torres, L., Miller, M. J., & Moore, K. M. (2013). Factorial invariance of the Brief Symptom Inventory-18 (BSI-18) for adults of

Mexican descent across nativity status, language format, and gender. *Psychological Assessment, 25*, 300–305.
Wang, J., Kelly, B. C., Booth, B. M., Falck, R. S., Leukefeld, C., & Carlson, R. G. (2010). Examining factorial structure and measurement invariance of the Brief Symptom Inventory (BSI)-18 among drug users. *Addictive Behaviors, 35*, 23–29. https://doi.org/10.1016/j.addbeh.2009.08.003
Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives, 4*, 10–18.
Wiesner, M., Chen, V., Windle, M., Elliott, M. N., Grunbaum, J. A., Kanouse, D. E., & Schuster, M. A. (2010). Factor structure and psychometric properties of the Brief Symptom Inventory-18 in women: A MACS approach to testing for invariance across racial/ethnic groups. *Psychological Assessment, 22*, 912–922.
Wu, P.-C. (2016). Response shifts in depression intervention for early adolescents: Response shifts in depression intervention. *Journal of Clinical Psychology, 72*, 663–675. https://doi.org/10.1002/jclp.22291
Zabora, J., Brintzenhofeszoc, K., Jacobsen, P., Curbow, B., Piantadosi, S., Hooker, C., ... Derogatis, L. (2001). A new psychosocial screening instrument for use with cancer patients. *Psychosomatics, 42*, 241–246.

Received September 3, 2017
Revision received January 1, 2018
Accepted January 8, 2018
Published online August 3, 2018
EJPA Section/Category Clinical Psychology

Gerrit Hirschfeld
Quantitative Methods
University of Applied Sciences Osnabrück
Caprivistr. 30A
49074 Osnabrück
Germany
g.hirschfeld@hs-osnabrueck.de

