

Psychometric Properties and Validation of the English Version Giessen Subjective Complaints List (GGB-8)

Katja Petrowski (✉ katja.petrowski@tu-dresden.de)

Dresden University of Technology

Markus Zenger

University of Applied Sciences Magdeburg-Stendal

Bjarne Schmalbach

University Medical Center of the Johannes Gutenberg University of Mainz

Christina Diane Bastianon

University Medical Center of the Johannes Gutenberg University of Mainz

Bernhard Strauss

University Hospital Jena

Research Article

Keywords: Giessen Subjective Complaint List, English version, psychometric properties, validation, assessment, psychosomatic

Posted Date: January 18th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1247630/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Psychology on March 11th, 2022. See the published version at <https://doi.org/10.1186/s40359-022-00741-8>.

Abstract

Background

The present study investigated the psychometric properties of the newly developed English version of the Giessen Subjective Complaint List – 8 (GGB-8), a questionnaire assessing psychosomatic symptoms with regard to exhaustion, gastrointestinal, musculoskeletal and cardiovascular.

Methods

A U.S. sample of 638 participants (47.6% female) was recruited by MTurk to participate in this cross-sectional online survey. Validation instruments included the Patient Health Questionnaire-4, Perceived Stress Scale, short version of the Trier Inventory for Chronic Stress.

Results

Reliability was high with ω 's between .80 and .86 for all subscales. Confirmatory factor analyses yielded comparable good model fit for a four-dimensional model as well as a higher order model. Multi-group confirmatory factor analyses confirmed measurement invariance of the GGB-8 across sex and age. Regarding convergent validity, correlations with other instruments were highly significant and of large magnitude as expected.

Conclusion

The English version of the GGB-8 has shown excellent psychometric properties. Therefore, it can be recommended for the assessment of psychosomatic complaints in contexts where short screening instruments are necessary.

Background

Somatic complaints are highly represented in society [1–3] as well as in patients involved in the health care system [4, 5]. Although these are not directly connected to medical conditions, patients with severe medical conditions commonly utter somatic symptoms. However, somatic symptoms are associated to symptoms of anxiety and depression [2]. The assessment of somatic symptom stress is key in epidemiological research, since such are known to reduce the health-related quality of life and are related to a greater use of the health care services [2, 6, 7].

So far, the assessment comparability of the somatic symptom burden in the general population has been limited by a lack of agreement on the scales used stated by Zijlema and colleagues [8]. In their recent systematic review, Zijlema and colleagues [8] identified 40 self-report somatic symptoms questionnaires

and assessed them regarding their usability for large scale population studies. The authors suggested the Patient Health Questionnaire-15 [9] and the somatization scale of Symptom Checklist 90 [10]. However, the Symptom Checklist 90 lacks the brevity needed in epidemiological research. Even for gerontic patients or older patients 15 items of the PHQ might be still too long. Therefore, shorter version of the PHQ (including the PHQ-4) and other scales with the assessment of somatic symptoms should be used. Shorter psychometrically sound questionnaires have several benefits: the drop-out rate as well as the rate of missing values are lower in shorter surveys, and the participants experience less boredom or fatigue.

In German speaking societies, a scientifically sound scale for the assessment of subjective health complaints is the Giessen Subjective Complaints List (Gießener Beschwerdebogen – GBB [11]). The GBB-24 consists of 24 health complaints rated on a Likert scale ranging from 0 (not at all) to 4 (very much). The individual complaints can be aggregated on four scales: exhaustion, gastrointestinal complaints, musculoskeletal complaints, cardiovascular complaints. These scales correspond well to commonly reported symptom clusters [8]. After continuous improvement, the 24-item version of the GBB-24 came into application for the evaluation of physical complaints after medical assessments, social stressors, psychotherapy, symptom strain in minority and marginalized groups. The GBB is also used for basic documentation in psychosomatic medicine and psychotherapy [11]. In order to use the GBB in epidemiological research, a shorter version would be necessary. Therefore, an 8-item brief version of the Giessen Subjective Complaints List was developed. The following criteria were applied for the shortened scale: (1) maintaining the original factor structure and having an equal number of items per factor (as is the case in the original long form), (2) the selected items should be among those with the highest item total-correlation from each subscale, (3) the selected items should have a mean above .5 in the general population to avoid floor effects.

Psychometric analyses of the German version of the GBB-8 yielded excellent scale properties with regard to item characteristics and factor structure. The eight symptoms included in the questionnaire are among the top 15 symptoms reported by Zijlema and colleagues as the most frequently assessed [8]. This shows the relevance of the chosen criteria not only due to their psychometric quality but also regarding their content. A factor structure that allows for the computation of subscales, including norms for each subscale, provides an advantage over measures providing only one overall score. Strong measurement invariance can be largely confirmed regarding gender, age, and age x gender. The factors are more easily interpretable and highlight the specific areas of complaint. Given the norms and the confirmed factor structure, the subscales can be used independently.

In sum, the psychometric properties of the GBB-8 are proven to be excellent. Therefore, to utilize this scale in different languages an English version was translated and the psychometric properties were tested in a native English-speaking population. In order to test the validity chronic stress was assessed since high chronic stress is accompanied with more severe somatic symptoms and lower chronic stress with less somatic symptoms.

Method

Participants

The sample is comprised of 638 participants. Males and females were relatively equally represented, 47.6% females and 52.2% males. The majority were aged less than 40 (66%) while 15% were 40-49 years old and 19% older than 49. Most participants reported being married (72.3%) and having a bachelor's degree education (63%). The participants self-identified as ethnically white (79.9%), black (8.6%), Hispanic (5.2%) and Asian (4.1%). A full overview of sample characteristics is found in Table 1.

Data source

Amazon Mechanical Turk (MTurk) was used to recruit participants living in the United States (U.S.) to complete an online survey, which covered topics of acute and chronic stress and uncertainty, physical complaints, emotion regulation, sleep and health behavior. The survey was implemented using SoSci Survey [12], a German based survey platform.

The survey was posted as a HIT (Human Intelligence Task) on MTurk in Fall 2020. The task asked that workers complete an externally hosted survey in exchange for \$0.50. The HIT was titled "20-25 min. Psychological Survey about Stress and Uncertainty" and described as "This survey aims to investigate stress and uncertainty during the COVID pandemic and validate a psychological scale with English speakers". The HIT was visible only to workers with an acceptance rate greater than 95% and who were residents in the U.S. To prevent workers from completing the HIT twice, a qualification was given to all workers that restricted them from partaking in the second round. After completing the survey, they were given an automatically generated code, which was required to provide in MTurk for payment (no workers were rejected for payment).

Several instructional manipulation checks were embedded in the survey as a response quality check [13, 14]. It asked participants to respond to a separate question using the same response options. If incorrect, they were warned to carefully read the instructions and given a second chance to correctly answer. Further, quality control measures were included such as, 2. response consistency between birthdate and age, 3. open response questions were manually coded in line with the Chmielewski article (14) and 4. time checks on the quickness to complete the questionnaire were all checked. Participants who failed to correctly respond after the warning were excluded from the analysis, which is the most effective method based on the literature (see 14). From the full sample, 28 observations were dropped due to complete missing data and an additional 205 participants were excluded based on the response quality checks. Therefore, the sample was reduced to N=638. According to one missing in the assessment of sex, the sample size was reduced to N=637 in cases where the variable sex was used for calculations.

Measures

The Gießen Subjective Complaints List (GBB-8, [15]) is a short and reliable instrument for evaluating the degree of somatic symptoms. The eight items identify commonly reported complaints whereby participants respond on 5-point Likert. The GBB-8 was translated into English in accordance with the International Test Commission (ITC) Guidelines for Translating and Adapting Tests [16]. The items were translated from German to English by one bilingual expert and then back-translated to German by a second bilingual expert. Comparison and reconciliation of the original and back-translated items was carried out by a group of experts, followed by a second round of forward and back-translation. The English GBB-8 items are included in Appendix A.

The Patient Health Questionnaire (PHQ-4, [1]) is a 4-item inventory to very briefly identify depression and anxiety. Items stem from the Generalized Anxiety Disorder (GAD-7) and the PHQ-8. Participants rate items on a 4-point Likert scale. The two-factor structure is represented by the two anxiety items (Factor 1) and the two depression items (Factor 2). The two factors explained 84% of the total variation and factor loadings were all $\geq .82$ [17]. Reliability of PHQ-4 scales are good (Cronbach $\alpha > 0.80$) [17].

The Perceived Stress Scale (PSS-10) [18] measures the degree to which life has been experienced as unpredictable, uncontrollable and overloaded over the past month. Participants respond on a 5-point Likert scale. Cohen et al. [18] originally developed the PSS as a single factor, however since its development, many researchers have concluded the scale represents two distinct factors: (1) perceived helplessness and (2) perceived self-efficacy [19–21]. The PSS-10 consistently shows strong internal reliability (Cronbach $\alpha > .70$) in diverse populations and meets the criteria for good test-retest validity ($> .70$) [22].

The short-English version of the Trier Inventory for Chronic Stress (TICS-9) is based on the original 57 item scale [23] that was translated into English, shortened and validated [24, 25]. The TICS-9 represents nine factors of chronic stress. These include: Work Overload; Social Overload; Pressure to Perform; Work Discontent; Excessive Demands at Work; Lack of Social Recognition; Social Tensions; Social Isolation; Chronic Worrying. Participants rate the frequency of specific situations over the previous three months on a 5-point Likert scale. The English TICS-9 reflects the strengths of the full 57-item English TICS [24]; as it is reliable (Cronbach $\alpha \geq .86$), shows good model fit, and the scale structure is invariant between males and females supporting the scale validity [25].

Statistical analyses

All analyses were conducted in R, using the packages lavaan, moments, multilevel, and semTools [26–29]. There was only a small amount of missing data (166 of 6744 GBB data points; i.e. 2.5%). Nonetheless, we ran confirmatory factor analysis using robust full-information maximum likelihood estimation to deal with missing values and non-normal distributions [30, 31]. For the evaluation of model fit we followed the guidelines provided by Schermelleh-Engel et al. [32]: a non-significant χ^2 , Comparative Fit Index / Tucker-Lewis Index (CFI/TLI) greater than .95 (.97), Root Mean Square Error of Approximation (RMSEA) smaller than .08 (.05), and Standardized Root Mean Square Residual (SRMR) smaller than .10

(.05) for acceptable (good) fit. For CFI, TLI, and RMSEA we used the robust variants [33, 34]. We report McDonald's ω as a reliability metric [35].

Next, we tested for measurement invariance by comparing CFI and RMSEA between models that did (did not) constrain the measurement parameters (loadings, intercepts, residuals) to be equal between the groups of interest [36]. Specifically, these include the configural (unconstrained), metric (loadings constrained), scalar (loadings and intercepts constrained), and the strict (loadings, intercepts, and residuals constrained) model. Δ CFI and Δ RMSEA should be smaller than .010 and .015, respectively. Since we incorporated a higher-order construct in our model, we followed the guidelines provided by Chen et al. [37] and tested first- and second-order invariance successively. After establishing strict invariance on both the first and second factor order, we then also examined the latent mean differences by additionally constraining the higher-order factor to be equivalent between groups. In addition to the χ^2 -test, we examined the standardized factor mean difference using the formula:

$$f = \frac{\sqrt{\frac{\sum_i^k (\alpha_i - \bar{\alpha})^2 * n_i}{n_{total}}}}{\psi_p} \quad (1),$$

with

$$\bar{\alpha} = \frac{\sum_i^k (\alpha_i * n_i)}{n_{total}} \quad (2),$$

where ψ_p is the standard deviation of the respective factor, pooled across all tested groups, n_k is the sample size of group k , and α_k is the latent variable mean in group k .

Results

Item descriptive statistics

In Table 2, we report descriptive item statistics. Skewness and kurtosis for all eight items indicate normal distribution [38]. In addition, we calculated the squared Mahalanobis distance and tested it for significance to identify outlier cases. A total of 4 (0.6%) of cases were flagged as outliers but were retained in the analysis. Removing these cases from the analyses did not meaningfully change the outcomes. The subscale-specific correlations as well as the item-total correlations for the total score were high. This was to be expected as the GBB-8 is a homogenous instrument that assesses a relatively narrow construct.

Factorial validity

We tested a total of three different model configurations. First, we tested a unifactorial model, which evinced acceptable fit, $\chi^2(20) = 133.636$, $p < .001$, CFI = .965, TLI = .951, RMSEA = .100, SRMR = .030. Model fit improved substantially by grouping the items onto their specific subscales in a four-dimensional model, indicating that the higher-order factor aligns well with the empirical data structure, $\chi^2(14) = 57.740$, $p < .001$, CFI = .988, TLI = .975, RMSEA = .071, SRMR = .018. Finally, we expanded the four-dimensional model by adding a second-order latent variable, representing general somatic symptom burden. This model had virtually the same fit as the four-dimensional one, showing that a general construct underlying the four subscales can be assumed, $\chi^2(16) = 64.529$, $p < .001$, CFI = .985, TLI = .974, RMSEA = .072, SRMR = .021. In terms of reliability, all subscales evinced good coefficients in the four-dimensional model with a second-order construct, with ω 's between .80 and .86. In addition, the reliability of the second order factor was excellent by all accounts: The vast majority of variance in both Level 1 ($\omega_{L1} = .922$) and Level 2 ($\omega_{L1} = .977$) is explained by the second order factor.

Measurement Invariance

We then tested the second-order factor model for measurement invariance across sex and age. As can be seen in Table 3, model fit decreases were negligible upon introducing the various constraints. Thus, the model can be assumed invariant across sex and age. As a result, comparisons of both, latent and observed means and variances, are admissible. Finally, we constrained the latent means of the second-order factor to be equal between groups to check whether there are significant differences between groups. Here it became clear that for sex there was virtually no difference ($p(\chi^2) = .169$, $d = 0.10$), whereas for age there were significant but small differences ($p(\chi^2) = .012$, $R^2 = .014$).

Convergent validity

In Table 4 we report correlations between the GBB subscales and total score and related scales of psychological symptoms. As expected, all correlations were highly significant and of large magnitude.

Normative Scores

Normative percentile ranks for the GBB subscale and total scores are reported in Table 5 and in Table 6.

Discussion

The assessment of somatic symptom stress is of high relevance in the context of epidemiological research [2, 6, 7]. In a recent systematic review [8] the Patient Health Questionnaire-15 [9] was primarily recommended as a measure of somatic symptom stress. Since epidemiological studies need short forms of measures, shorter psychometrically sound questionnaires might have several benefits. The short versions of the PHQ, e.g., PHQ-4, do not include items to assess somatic symptom burden. Therefore, the

German GBB-8, established in German speaking countries, was translated into English and its psychometric properties were assessed.

The present study showed that the four specific subscales fit well with the four-dimensional empirical model which indicates a good factorial fit of the GBB-8. In addition, the higher-order factor representing general somatic symptom burden aligns well with the empirical data structure as well. It is impressive that this higher-order model had virtually the same fit as the four-dimensional one. Therefore, it can be assumed that a general construct underlies the four subscales. Thus, the model can be assumed invariant across sex and age. As a result, comparisons of both, latent and observed means and variances, are admissible. Finally, we constrained the latent means of the second-order factor to be equal between groups to check whether there are significant differences between groups. Here it became clear that for sex there was virtually no mean difference, whereas for age there were significant but very small differences between the models which can be neglected. Age has a well-known effect on somatic burden. In this respect it is noteworthy that the majority of age of the sample is younger than 40 (66%) and only 19% older than 49. The older age with the higher somatic symptom burden is not present in the sample which limits the generalizability of the age invariance. The psychometric properties of the GBB have to be reevaluated in a larger representative sample with better distribution of age. Furthermore, sampling with MTurk lead to a larger percentage of highly educated (63%) and more married participants (72.3%). The U.S. MTurk population is known to be younger and more educated in comparison to U.S. representative samples [39, 40], this divergence is common among internet users in general and other online survey methods [40]. However, in line with the younger age, MTurk users are less likely to be married[40], which is inconsistent with the current sample. This may be a result of when the samples were collected (e.g. 2016 and 2020), however the U.S. Census Bureau reports a decline in marriage rates from 2009 to 2019 [41]. Marital status and education influence the somatic symptom burden (Petrowski et al. 2015) and might therefore reduce the generalizability of the current results.

Concerning reliability, all subscales showed good reliability even for the second-order construct. The reliability of the second order factor was excellent. Concerning the validity correlations between the GBB subscales and related scales of psychological symptoms were highly significant and of large magnitude. Therefore, a convergent validity is given and proves the excellent psychometric properties of this short instrument.

Strengths & Limitations

The strength of the English version of the GBB-8 is the excellent psychometric properties with the good factorial structure, the convergent validity and the brevity of the inventory showing its usefulness especially for epidemiological surveys. The study also has a few limitations: First, the data was based on a MTurk sampling, which conveys some limitations in generalizability. The MTurk population is not representative of the U.S. population, however studies have shown MTurk samples are comparable to other traditional subject pools [42]. Second, while Instruction Manipulation Checks (IMC) are shown to

reduce sample noise due to non-diligent participants (including bots and farmers) [13], by dropping participants who failed the IMC, there is the potential to harm the external validity of the study [13].

Conclusion

The GBB-8 is a carefully designed instrument that possesses good psychometric properties. In addition, the applicability of the GBB-8 in different subpopulations is a unique characteristic of this instrument. Now it is even applicable in English speaking surveys.

List Of Abbreviations

GBB- Giessen Subjective Complaints List [Gießener Beschwerdebogen]

MTurk- Amazon Mechanical Turk

U.S.- United States

HIT- Human Intelligence Task

ITC- International Test Commission

PHQ- Patient Health Questionnaire

GAD- Generalized Anxiety Disorder

PSS- Perceived Stress Scale

TICS- Trier Inventory of Chronic Stress

CFI- Comparative Fit Index

TLI- Tucker Lewis Index

RMSEA- Root Mean Square Error of Approximation

SRMR- Standardized Root Mean Square Residual

IMC- Instructional Manipulation Check

Declarations

Ethics approval and consent to participate

All participants were provided a data protection declaration and informed of the aims of the study. They provided written informed consent in lines with the Helsinki Declaration.

The study was approved by the Ethical commission of Rheinland-Pfalz State Medical Association [Landesärztekammer Rheinland-Pfalz] (2019-14290) and conducted in accordance to the ethical guidelines of the Ethical commission of Rheinland-Pfalz State Medical Association

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analyzed during the current study are available from Prof. Katja Petrowski: kpetrows@uni-mainz.de on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

KP and MZ supervised project and data collection. BS conducted statistical analyses. CB developed and managed Mturk survey and collected data. KP, MZ, BS, CB, and BHS all contributed equally to the manuscript and provided valuable feedback in its development. All authors read and approved the final manuscript.

Acknowledgments

The authors are grateful for the participants' contribution to this research.

References

1. Kroenke K, Arrington M., Mangelsdorff A. The prevalence of symptoms in medical outpatients and the adequacy of therapy. *Arch Intern Med.* 1990;150.
2. Gierk B, Kohlmann S, Kroenke K, Spangenberg L, Zenger M, Brähler E, et al. The Somatic Symptom Scale–8 (SSS-8). *JAMA Intern Med.* 2014;174:399. doi:10.1001/jamainternmed.2013.12179.
3. Creed FH, Davies I, Jackson J, Littlewood A, Chew-Graham C, Tomenson B, et al. The epidemiology of multiple somatic symptoms. *J Psychosom Res.* 2012;72:311–7. doi:10.1016/j.jpsychores.2012.01.009.
4. Löwe B, Spitzer RL, Williams JBW, Mussell M, Schellberg D, Kroenke K. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *Gen Hosp Psychiatry.*

- 2008;30:191–9. doi:10.1016/j.genhosppsy.2008.01.001.
5. Kroenke K, Spitzer RL, Williams JB, Linzer M, Hahn SR, de Gruy III FV, et al. Physical symptoms in primary care. Predictors of psychiatric disorders and functional impairment. *Arch Fam Med*. 1994;3:774–9. doi:10.1001/archfami.3.9.774.
 6. Kroenke K, Zhong X, Theobald D, Wu J, Tu W, Carpenter JS. Somatic Symptoms in Patients With Cancer Experiencing Pain or Depression. *Arch Intern Med*. 2010;170. doi:10.1001/archinternmed.2010.337.
 7. Kohlmann S, Gierk B, Hümmelgen M, Blankenberg S, Löwe B. Somatic Symptoms in Patients With Coronary Heart Disease: Prevalence, Risk Factors, and Quality of Life. *JAMA Intern Med*. 2013;173:1469. doi:10.1001/jamainternmed.2013.6835.
 8. Zijlema WL, Stolk RP, Löwe B, Rief W, White PD, Rosmalen JGM. How to assess common somatic symptoms in large-scale studies: A systematic review of questionnaires. *J Psychosom Res*. 2013;74:459–68. doi:10.1016/j.jpsychores.2013.03.093.
 9. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: Validity of a New Measure for Evaluating the Severity of Somatic Symptoms. *Psychosom Med*. 2002;64:258–66.
 10. Derogatis LR, Unger R. Symptom Checklist-90-Revised. In: *The Corsini Encyclopedia of Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010. doi:10.1002/9780470479216.corpsy0970.
 11. Brähler E, Hinz A, Scheer JW. GBB-24. Der Gießener Beschwerdebogen Manual. 3rd edition. Bern: Huber; 2008.
 12. Leiner DJ. SoSci Survey. 2019. <https://www.socisurvey.de>.
 13. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: Detecting satisficing to increase statistical power. *J Exp Soc Psychol*. 2009;45:867–72. doi:10.1016/j.jesp.2009.03.009.
 14. Chmielewski M, Kucker SC. An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results. *Soc Psychol Personal Sci*. 2019. doi:10.1177/1948550619875149.
 15. Kliem S, Lohmann A, Klatt T, Mößle T, Rehbein F, Hinz A, et al. Brief assessment of subjective health complaints: Development, validation and population norms of a brief form of the Giessen Subjective Complaints List (GBB-8). *J Psychosom Res*. 2017;95:33–43. doi:10.1016/j.jpsychores.2017.02.003.
 16. International Test Commission. Guidelines for Translating and Adapting Tests. 2005. .
 17. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ–4. *Psychosomatics*. 2009;50:613–21.
 18. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav*. 1983;24:385–96. doi:10.2307/2136404.
 19. Hewitt PL, Flett GL, Mosher SW. The Perceived Stress Scale: Factor structure and relation to depression symptoms in a psychiatric sample. *J Psychopathol Behav Assess*. 1992;14:247–57. doi:10.1007/BF00962631.
 20. Roberti JW, Harrington LN, Storch EA. Further Psychometric Support for the 10-Item Version of the Perceived Stress Scale. 2006. doi:10.1002/j.2161-1882.2006.tb00100.x.

21. Bastianon CD, Klein EM, Tibubos AN, Brähler E, Beutel ME, Petrowski K. Perceived Stress Scale (PSS-10) psychometric properties in migrants and native Germans. *BMC Psychiatry*. 2020;20. doi:10.1186/s12888-020-02851-2.
22. Lee E-H. Review of the Psychometric Evidence of the Perceived Stress Scale. *Asian Nurs Res (Korean Soc Nurs Sci)*. 2012;6:121–7. doi:10.1016/j.anr.2012.08.004.
23. Schulz P, Schlotz W. Trierer Inventar zur Erfassung von chronischem Streß (TICS): Skalenkonstruktion, teststatistische Überprüfung und Validierung der Skala Arbeitsüberlastung. *Diagnostica*. 1999;45:8–19. doi:10.1026//0012-1924.45.1.8.
24. Petrowski K, Kliem S, Sadler M, Meuret AE, Ritz T, Brähler E. Factor structure and psychometric properties of the english version of the trier inventory for chronic stress (TICS-E). *BMC Med Res Methodol*. 2018;18. doi:10.1186/s12874-018-0471-4.
25. Petrowski K, Braehler E, Schmalbach B, Hinz A, Bastianon CD, Ritz T. Psychometric Properties of an English Short Version of the Trier Inventory for Chronic Stress. *BMC Med Res Methodol*. 2020. doi:10.21203/rs.2.22524/v1.
26. Bliese P. multilevel: Multilevel Functions. 2016. <https://cran.r-project.org/package=multilevel>.
27. Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y. semTools: Useful tools for structural equation modeling. 2019.
28. Komsta L, Novomestky F. moments: Moments, cumulants, skewness, kurtosis and related tests. 2015. <https://cran.r-project.org/package=moments>.
29. Rosseel Y. lavaan: An R package for Structural Equation Modeling. *J Stat Softw*. 2012;48:1–36. <http://www.jstatsoft.org/v48/i02/>.
30. Enders C, Bandalos D. The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Struct Equ Model A Multidiscip J*. 2001;8:430–57. doi:10.1207/S15328007SEM0803_5.
31. Yuan K-H, Bentler PM. 5. Three Likelihood-Based Methods for Mean and Covariance Structure Analysis with Nonnormal Missing Data. *Sociol Methodol*. 2000;30:165–200. doi:10.1111/0081-1750.00078.
32. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods Psychol Res Online*. 2003;8:23–74. https://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf. Accessed 8 Feb 2019.
33. Brosseau-Liard PE, Savalei V. Adjusting Incremental Fit Indices for Nonnormality. *Multivariate Behav Res*. 2014;49:460–70.
34. Brosseau-Liard PE, Savalei V, Li L. An Investigation of the Sample Performance of Two Nonnormality Corrections for RMSEA. *Multivariate Behav Res*. 2012;47:904–30.
35. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol*. 2014;105:399–412. doi:10.1111/bjop.12046.

36. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Model*. 2007;14:464–504.
37. Chen FF, Sousa KH, West SG. Teacher's Corner: Testing Measurement Invariance of Second-Order Factor Models. *Struct Equ Model A Multidiscip J*. 2005;12:471–92. doi:10.1207/s15328007sem1203_7.
38. Kim H-Y. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. 2013;38:52. doi:10.5395/rde.2013.38.1.52.
39. Paolacci G, Chandler J. Inside the Turk. *Curr Dir Psychol Sci*. 2014;23:184–8. doi:10.1177/0963721414531598.
40. Chandler J, Shapiro D. Conducting Clinical Research Using Crowdsourced Convenience Samples. *Annu Rev Clin Psychol*. 2016;;53–81. doi:10.1146/annurev-clinpsy-021815-093623.
41. Anderson L, Scherer Z. U.S. Marriage and Divorce Rates Declined in Last 10 Years. 2020. <https://www.census.gov/library/stories/2020/12/united-states-marriage-and-divorce-rates-declined-last-10-years.html>.
42. Paolacci G, Chandler J, Ipeirotis PG, Stern LN. Running experiments on Amazon Mechanical Turk. 2010. <http://sjdm.org/~baron/journal/10/10630a/jdm10630a.pdf>. Accessed 28 Jan 2019.

Tables

Table 1 *Sociodemographic sample details*

	<i>n</i>	%
Sex		
male	333	52.3
female	304	47.6
Missing	1	
Age group		
Younger than 30	239	37.5
30-39	182	28.5
40-49	96	15
Older than 49	121	19
Marital status		
Single	115	18
Married	461	72.3
Domestic partnership	30	4.7
Separated	21	3.3
Divorced	6	0.9
Widowed	3	0.5
Ethnicity		
Hispanic or Latino	33	5.2
Asian	26	4.1
Black or African American	55	8.6
Native Hawaiian or Other Pacific Islander	2	0.3
White	510	79.9
Multi-ethnic	5	0.8
Preferred not to answer	5	0.8
Education		
High School	50	7.8
Associate's degree	43	6.7
Bachelor's degree	402	63

Masters's degree	121	19
Professional degree	8	1.3
Doctorate degree	8	1.3
Other	4	0.6
Household income		
\$0-\$1249	168	26.3
\$1250-\$3499	219	34.3
\$3500-\$4999	99	15.5
\$5000 and more	142	22.3
Preferred not to answer	10	1.6

Table 2 *Descriptive item statistics*

	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	r_{subscale}	r_{it}
Being easily exhausted ^{Ex}	1.73	1.36	0.06	1.71	.66	.80
Tiredness ^{Ex}	2.06	1.25	-0.19	2.01	.66	.74
Feeling bloated or distended ^{Gast}	1.71	1.36	0.10	1.76	.75	.81
Stomachache ^{Gast}	1.85	1.38	0	1.72	.75	.81
Backache ^{Musc}	1.98	1.30	-0.12	1.90	.68	.73
Neck or shoulder pain ^{Musc}	1.95	1.31	-0.09	1.88	.68	.73
Palpitations or heart pounding ^{Card}	1.53	1.28	0.21	1.83	.68	.76
Dizziness ^{Card}	1.7	1.34	0.09	1.74	.68	.79

Note: *M* = mean, *SD* = standard deviation, r_{subscale} = subscale-specific correlations, r_{it} = item-total correlations. Subscales: ^{Ex} = Exhaustion; ^{Gast} = Gastrointestinal; ^{Musc} = Musculoskeletal; ^{Card} = Cardiovascular.

Table 3 *Test of measurement invariance across sex and age*

Participant sex (male, female)

	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	<i>p</i>	<i>CFI</i>	ΔCFI	<i>RMSEA</i>	$\Delta RMSEA$
Configural model	81.981	32				.985		.072	
1 st order weak invariance	89.422	36	7.441	4	.114	.985	0	.069	.003
2 nd order weak invariance	93.273	39	3.851	3	.278	.985	0	.066	.003
1 st order strong invariance	99.694	43	6.421	4	.17	.985	0	.063	.003
2 nd order strong invariance	107.529	46	7.835	3	.05	.984	.001	.063	0
1 st order strict invariance	117.238	54	9.709	8	.286	.984	0	.059	.004
2 nd order strict invariance	126.662	58	9.424	4	.051	.982	.002	.06	-.001
Mean equivalence	128.558	59	1.896	1	.169	.982	0	.06	0

Age groups (<30, 30-39, 40-49, >49)

	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	<i>p</i>	<i>CFI</i>	ΔCFI	<i>RMSEA</i>	$\Delta RMSEA$
Configural model	108.741	64				.987		.068	
1 st order weak invariance	124.371	76	15.63	12	.209	.987	0	.062	.006
2 nd order weak invariance	135.123	85	10.752	9	.293	.987	0	.059	.003
1 st order strong invariance	145.451	97	10.328	12	.587	.988	-.001	.053	.006
2 nd order strong invariance	153.651	106	8.2	9	.514	.988	0	.05	.003
1 st order strict invariance	194.985	130	41.334	24	.015	.984	.004	.053	-.003
2 nd order strict invariance	202.323	142	7.338	12	.834	.984	0	.05	.003
Mean equivalence	213.344	145	11.021	3	.012	.982	.002	.052	-.002

Note: χ^2 = chi square, *df*= degrees of freedom, Δ = change from previous model, *p* = significance level, CFI = Comparative Fit Index, RMSEA= Root Mean Square Error of Approximation.

Table 4 *Correlation matrix*

	1	2	3	4	5	6	7	8	9
1 - GBB-Exhaustion	-	.79	.73	.77	.91	.65	.66	.75	.65
2 - GBB-Gastrointestinal		-	.71	.81	.92	.61	.61	.73	.63
3 - GBB-Musculoskeletal			-	.69	.87	.61	.61	.66	.6
4 - GBB-Cardiovascular				-	.91	.59	.59	.72	.63
5 - GBB-Total					-	.68	.69	.79	.7
6 - PHQ-Anxiety						-	.71	.75	.68
7 - PHQ-Depression							-	.73	.62
8 - TICS								-	.81
9 - PSS									-

Table 5 *Normative percentile ranks for the GBB subscale scores*

Sum Score	Exhaustion	Gastrointestinal	Musculoskeletal	Cardiovascular
0	14	20	14	23
1	23	29	21	31
2	32	37	29	40
3	44	44	40	48
4	54	59	53	67
5	73	73	72	77
6	84	87	81	94
7	98	94	97	98
8	100	100	100	100

Table 6 *Normative percentile ranks for the GBB total scores*

Sum Score	Percentile Rank	Sum Score	Percentile Rank	Sum Score	Percentile Rank
0	8	11	36	22	79
1	11	12	38	23	83
2	14	13	40	24	87
3	18	14	44	25	90
4	21	15	46	26	92
5	23	16	52	27	94
6	25	17	55	28	98
7	28	18	62	29	100
8	31	19	66	30	100
9	32	20	72	31	100
10	34	21	76	32	100

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixGBB.docx](#)