

Psychometric Properties of an English Short Version of the Trier Inventory for Chronic Stress

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

Technische Universität Dresden <https://orcid.org/0000-0001-7297-2093>

Elmaer Braehler

Department of Psychosomatic Medicine and Psychotherapy, University Medical Center of the Johannes Gutenberg University-Mainz

Bjarne Schmalbach

Medical Psychology and Medical Sociology, University Medical Center of the Johannes Gutenberg University Mainz

Andreas Hinz

Department of Medical Psychology and Medical Sociology, University of Leipzig

Christina Bastianon

Johannes Gutenberg Universität Mainz

Thomas Ritz

Department of Psychology, Southern Methodist University

Research article

Keywords: Chronic stress, screening questionnaire, factor analysis, measurement invariance, Trier Inventory for Chronic Stress

Posted Date: November 9th, 2020

DOI: <https://doi.org/10.21203/rs.2.22524/v3>

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 on December 16th, 2020. See the published version at <https://doi.org/10.1186/s12874-020-01156-y>.

Abstract

Background: Although a variety of instruments are available that capture stress experience, the assessment of chronic stress has been hindered by the lack of economical screening instruments. Recently, an English-language version of the Trier Inventory for Chronic Stress (TICS-EN) consisting of 57 items according to a systemic-requirement-resource model of health in nine subdomains of the chronic stress experience has been introduced.

Methods: We constructed a new 9-item short version of the TICS covering all nine subdomains and evaluated it in two samples (total $N=685$). We then used confirmatory factor analysis to check factorial validity.

Results: This version showed a highly satisfactory model fit, was invariant across participant gender, demonstrated a very high correlation with the original TICS ($r=.94$), and showed a moderate correlation ($r=.58$) with a measure of perceived stress in the past month.

Conclusions: Therefore, this theoretically driven instrument can be recommended as a short version of the TICS in English language.

Background

According to the Federal Institute for Occupational Safety and Health, the levels of perceived stress have significantly increased since the early 2000s [1]. Consequences of long-lasting chronic stress constitute an increased risk for impaired psychological wellbeing and acute physical illnesses [2,3]. Associations between psychosocial stress and depression, cardiovascular disease, sleep disorders or chronic pain are well-established [4–9]. Elevated stress levels are also a factor in susceptibility to upper respiratory tract infections, asthma, herpes viral infections, autoimmune diseases, and delayed wound-healing [9]. Accordingly, measuring chronic stress with a brief and efficient instrument is valuable to multiple disciplines [10].

An in-depth overview of general- and area-specific stress instruments is provided by Cohen and colleagues [3]. However, the focus on acute stress has often overlooked chronic, long-term stress. For chronic stress, the Perceived Stress Questionnaire [11,12] provides a general retrospective one-year evaluation in addition to the last four, six, and eight weeks without area specificity. The Trier Inventory for Chronic Stress (TICS) [13] is an instrument that targets explicitly area-specific chronic stress including Work Overload, Social Overload, Pressure to Perform, Work Discontent, Excessive Demands from Work, Lack of Social Recognition, Social Tensions, Social Isolation, and Chronic Worrying. These nine domains of chronic stress are measured by 57 items [13,14], which were selected in accordance with the systemic-requirement-resource model of health [15]. Schulz, Schlotz, and Becker [13] developed the TICS scales based on this model and thus assumed high content validity as a rational response. In a representative German study the confirmatory factor analysis (CFA) provided evidence for a good factorial validity [14].

A 12-item version of the TICS (Short Screening Scale for Chronic Stress, SSCS) was developed by the original authors [13] to meet the needs for a brief chronic stress instrument for applied research and practitioners. Items were selected based on factor loadings of the strong, unrotated first factor (explained variance 28.4%). However, this empirical item selection included only five of the original nine stress domains from the full 57-item scale in the SSCS. With the stress domains represented in the SSCS (Chronic Worry, Work Overload, Excessive Demands from Work, and Lack of Social Recognition) the SSCS correlated moderately to highly with the original nine subscales of the 57-item ($r = .68-.87$). Social Overload retained one item for the SSCS but correlated low with the subscale of the long version of the TICS ($r = .45$) [13]. The four scales not represented in the SSCS, Pressure to Perform, Work Discontent, Social Tensions, and Social Isolation, showed low correlations as well with the 12-item SSCS ($r = .40-.56$) [13]. Therefore, four of the nine theoretically proposed areas of chronic stress are underrepresented in the SSCS. Due to the item selection procedure, the content domain was reduced and thus the validity of the SSCS was weakened [16]. The strength and uniqueness of the original 57-item TICS lies in its area-specific chronic stress assessment based on the theoretical model of health, unfortunately this strength is no longer represented in the short 12-item version (TICS-12, SSCS).

In order to represent the theoretically and empirically supported nine areas of chronic stress, a new short version of the TICS was identified. Petrowski et al.[17]. A representative German sample of $N = 2,473$ was used to construct and test the new 9-item TICS. Nine items based on the alphamax algorithm were chosen to represent the nine areas of chronic stress from the original TICS (TICS-57). The one-factor-model of TICS-9 provided a good fit for the latent construct and showed good internal consistency ($\alpha = .88$). As the original 57-item and 9-items TICS were developed and tested in German, we sought to replicate the findings with the English TICS-EN version [18] and evaluate the psychometric properties of an English-language short form TICS-9.

Method

Participants and procedure

The data of Sample 1 was collected at two college campuses in the Eastern and Southwestern region of the USA. Participants were undergraduate introductory psychology students who contributed in return for course credit. The final pooled sample consisted of $n = 501$. They received a data protection declaration that is in agreement with the Helsinki Declaration. The study was approved by the institutional review boards of the involved university institutions and all participants provided written informed consent.

Participants of Sample 2 were recruited in Spring 2019 via Amazon's Mechanical Turk [19], a crowdsourcing website. MTurk is an international online platform that allows researchers to post tasks or questionnaires that participants complete in return for payment. In the current study, participants signed up via MTurk and were then directed to the online survey to complete the questionnaire. This survey was only available to participants who were located in the USA and their MTurk approval rating greater than 95%. The questionnaire took approximately 5 min to complete and participants were compensated \$0.50

for their time. Sample 2 was collected in order to evaluate the factorial structure in a study. The final sample consisted of N = 184 participants.

The study was approved by the ethic review boards of Landesärztekammer Rheinland-Pfalz, Germany, and all participants provided informed consent online by agreeing to take part in the study.

Table 1 *Sample description*

	Sample 1 (n = 501)			Sample 2 (n = 184)		
	n	%	TICS M (SD)	n	%	TICS M (SD)
Gender						
Female	366	73.1	2.47 (0.72)	96	52.2	2.44 (0.85)
Male	118	23.6	2.54 (0.70)	88	47.8	2.46 (0.89)
Missing	17	3.4	2.05 (0.84)			
Age (in years)	<i>M</i> = 19.97, <i>SD</i> = 2.84			<i>M</i> = 37.86, <i>SD</i> = 11.35		
≤20	366	73.1	2.50 (0.70)	2	1.1	3.44 (0.31)
21-25	116	23.2	2.42 (0.79)	16	8.7	2.78 (0.79)
≥26	19	3.8	2.37 (0.72)	166	90.2	2.41 (0.87)
Ethnicity						
White	391	78.0	2.49 (0.67)	148	80.4	2.35 (0.81)
Black or African American	22	4.4	2.58 (0.69)	16	8.7	2.26 (0.81)
Asian or Pacific Islander	25	5.0	2.28 (0.81)	6	3.3	2.41 (0.87)
Hispanic or Latino	6	1.2	2.74 (0.44)	13	7.1	2.94 (0.98)
Multi-ethnic	6	1.2	3.67 (0.78)	3	1.6	2.70 (0.39)
Other	25	5.0	2.12 (0.97)			
Missing	26	5.2	2.34 (0.91)	2	1.1	2.67 (0.63)

Note. TICS = Trier Inventory for Chronic Stress

Measures

The Trier Inventory for Chronic Stress (TICS) is a standardized German questionnaire that has been tested with respect to its factorial structure and psychometric properties, showing good to very good reliability [14]. Internal consistency (Cronbach's Alpha, α) was good to very good with values ranging from .84 to .91 (mean of α = .87) [13]. Nine interrelated factors of chronic stress are assessed: Work Overload; Social Overload; Pressure to Perform; Work Discontent; Excessive Demands at Work; Lack of Social Recognition; Social Tensions; Social Isolation; Chronic Worrying. The nine factors were derived from 57 items rated on a five-point rating scale (1-5, labeled as: "never", "rarely", "sometimes", "frequently", "always"). Participants rated the occurrence/frequency of specific situations with a recall period of the previous three months. The 12 items with the highest loadings constitute the short version by the original authors [13]. In addition, a new short version of the TICS was developed based on the alphamax algorithm representing the nine areas of chronic stress of the original TICS. The one-factor-model of this new short version provided a good fit for the latent construct and showed good reliability (α = .88) [17]. After the translation state-of-the-art (see Petrowski et al.[18]), the English version of the Trier Inventory of Chronic Stress (TICS-EN) with 57 items was used in the present study [18].

The Perceived Stress Scale (PSS-10) is the most widely used psychological instrument for measuring perceived stress [20]. Respondents report the degree to which situations in one's life have been unpredictable, uncontrollable and overloaded in the past month on a 5-point scale (0=never, 1=almost never, 2=sometimes, 3=fairly often, 4=very often).

Statistical Analyses

We conducted the analyses in R, using the packages *lavaan*, *lordif*, *MBESS*, and *semTools* [21–24]. Participants with missing values on any of the TICS-9 items were excluded from the analysis: seven and eleven participants. In addition, we excluded participants who failed the attention checks utilized in Sample 2 (n = 28). Very few participants (less than 5% across all items in both samples) chose the highest response option, making the items essentially ordinally scaled. Previous research suggests that conventional maximum likelihood estimation tends to be inaccurate with four or fewer response categories [25,26]. Therefore, we used the robust diagonally weighted least squares estimation method [27].

To evaluate model fit we considered the following measures and cutoff values[28–30]: The χ^2 -statistic should ideally be non-significant and is calculated by χ^2 divided by the degrees of freedom of the model, which should < than 2 to indicate good, or < 3 to indicate acceptable, fit. The Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI), which should be > .95 for good, or > .90 for acceptable fit, and finally, the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR), which should be lower than .08 to indicate acceptable, or <.05 to indicate good fit. Additionally, we report the 90% confidence interval for the RMSEA. In line with Dunn et al. [31], we tested reliability using McDonald's ω [32], accompanied by a 95% confidence interval.

In order to test for measurement invariance across gender groups, we used the approach of comparing increasingly constrained models as described by Milfont et al. [33]. Since we were dealing with ordered categorical data we modified the procedure in the way described by Wu et al.[34]: First, we compared the unconstrained (configural) model to a model with item thresholds fixed to be equal across groups. Second, the threshold-invariant model was compared to the metric model (item loadings constrained). Third, the metric model was compared to the scalar model (item intercepts constrained). To evaluate the model comparisons, we primarily used the differences in CFI and gamma hat (GH) between models – which should not exceed .01. Additionally, we tested for significant differences in χ^2 . To avoid selecting a non-invariant marker variable we estimated all factor loadings freely and set the variance of the latent variable to 1 instead. In addition, we analyzed differential item functioning using a logistic ordinal regression framework to be able to pinpoint the origin of whatever instances of measurement non-invariance we encountered [35–37].

Results

Item descriptive statistics

We report descriptive statistics for the TICS-9 items in Table 2 and Figure 1 and mean scores by sociodemographic group membership in Table 1. All corrected item-total correlations were above the commonly used cutoff value of .50.

Table 2 *Descriptive statistics for the TICS-9 (Sample 1)*

Item	<i>M</i>	<i>SD</i>	Skewness	Excessive Kurtosis	r_{it}	<i>CDIF</i>	<i>NCDIF</i>
1	1.838	0.910	0.949	0.467	.537	0.0117	0.0116
2	2.319	1.003	0.414	-0.306	.650	-0.0559	0.0356
3	3.275	1.066	-0.318	-0.358	.543	0.1627	0.1067
4	2.385	1.061	0.475	-0.444	.674	0.0288	0.0066
5	2.218	1.042	0.492	-0.504	.651	0.0082	0.0040
6	2.154	0.993	0.512	-0.510	.608	-0.0222	0.0029
7	2.669	1.184	0.236	-0.787	.594	0.1162	0.0771
8	2.375	1.001	0.406	-0.286	.573	-0.0276	0.0221
9	3.024	1.075	-0.193	-0.508	.580	0.0394	0.0062

Note. r_{it} = Corrected item-total correlation; *CDIF* = Compensatory differential item functioning ; *NCDIF* = Non-compensatory differential item functioning

Please insert Figure 1

CFA and reliability analysis

We report the results of the CFA in Table 3. Model fit was acceptable in both samples: Only the χ^2 -test indicated a significant deviation from the theoretical model. All fit indices presented evidence for acceptable, even good model fit. Internal consistency was satisfactory in both samples: $\omega_{\text{Sample1}} = .868$ [.850; .887], $\omega_{\text{Sample2}} = .872$ [.816; .927]. In comparison, the 57-item long version, which was included in Sample 1, had a substantially higher reliability coefficient, $\omega = .969$ [.965; .974].

Table 3 *Model fit in both samples*

	χ^2 (<i>df</i>)	<i>p</i>	χ^2/df	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i> [90% <i>CI</i>]	<i>SRMR</i>
Sample 1	116.878 (27)	< .001	4.329	.988	.984	.082 [.067; .097]	.060
Sample 2	40.614 (27)	.045	1.504	.997	.995	.073 [.011; .117]	.057

Measurement invariance

Next, we tested for measurement invariance across gender using Sample 1. Since not all items in all groups had sufficient frequencies for all response options, we collapsed the two highest replies “4” and “5”, putting the items on a four-point scale for the analysis of invariance. Utilizing the procedure described above we found evidence for invariance across gender. Only the χ^2 statistic showed a significant deviation, which was limited to the final comparison. CFI and GH never exceeded the cutoff of .01 between levels of constraints, indicating that women and men do not differ meaningfully in terms of their response behavior to the TICS-9.

When considering the indices of non-compensatory differential item functioning (NCDIF) presented in Table 2, it becomes clear that most of the gender-specific differences are attributable to Items 3 and 7, with both of them exceeding the cutoff of $NCDIF \leq 0.054$ for four-point items[38]. Thus, item-specific comparisons – specifically for these two items – are discouraged. However, considering the entire scale, differential test functioning (the sum of compensatory differential item functioning (CDIF), which accounts for the different directions of DIF; DTF = 0.2614), was below the critical value of $9 * .054 = 0.486$. Thus, overall there was a slight trend for women to choose higher response options than men, given the same trait level of stress (see Figure 2). However, this difference was so small that it was unlikely to meaningfully influence interpretation of test scores. This was also evident from Figure 2, which shows response behavior on the scale level across trait values.

Table 4 Analysis of measurement invariance across gender (Sample 1)

Model	χ^2 (df)	$\Delta\chi^2$	Δdf	p	CFI	ΔCFI	GH	ΔGH
Configural inv.	327.328 (56)				.956		.889	
Male	77.251 (28)				.966		.915	
Female	250.077 (28)				.952		.881	
Threshold inv.	340.100 (65)	12.772	9	.173	.955	.001	.888	.001
Loading inv.	346.736 (72)	6.636	7	.468	.955	.000	.888	.000
Intercept inv.	378.449 (80)	31.713	8	< .001	.951	.004	.879	.009

Please insert Figure 2.

Validity

In Sample 1, we found a very high correlation between the TICS-9 and the TICS-57, $r(499) = .944, p < .001$. Additionally, in Sample 2, there was a moderately high association of TICS-9 with PSS-10, $r_{\text{Sample2}}(182) = .583, p < .001$.

Discussion

A reliable nine-item version of the English TICS-EN was constructed for the assessment of chronic stress. The TICS-9-EN captures all areas of the systemic-requirement-resource model of health. Furthermore, the approach used to develop the English version TICS-9 benefited from the recommendations of Smith et al. [39], thus avoiding common methodological sins in developing short form scales [13,15,39]. Scales developed using only those items with the highest item-total correlations for a given factor [39–42] may uphold a high internal consistency estimate of reliability, but the content domain is unintentionally constricted and the validity of the short form is diminished [39,42]. For that reason, the alphamax algorithm was used to maximize reliability while maintaining all nine domains of chronic stress, thus avoiding this common sin. In order to avoid the challenge of impaired content coverage for a construct and inadequate validity from fewer items, TICS-9-EN covered all nine domains of the systemic-requirement-resource model of health [15] with one item each. The strength and uniqueness of the original TICS-57 was based on the theoretical model of health [13,15] and by maintaining content coverage, this advantage was also conveyed to the TICS-9-EN.

From a practical perspective, our short form combines the validity of the full-length version with an efficient version that is particularly suitable for large multivariable studies. It demonstrates a very good factorial structure and is highly correlated with the 57-item version. Meaningful interpretations can be made based on gender and age differences due to the invariance of the scale. Furthermore, the psychometric properties of the TICS-9-EN are similar to the original German long as well as short version of the TICS [17,18]: Model fit is good, so is reliability, and the scale can be considered invariant across pertinent sociodemographic groups. Future studies could complement the existing analyses by investigating cultural invariance between the two versions of the instrument.

The study has the strength that separate samples were drawn in order to replicate the factorial structure. However, some limitations need to be acknowledged. While the item selection algorithm maximizes Cronbach's alpha, at the same time, it increases model fit specifically for a one-factor solution. The improved psychometric properties of the new TICS-9 compared to the SCS is, consequently, partially a result of statistical methods, see Petrowski et al. [17]. Another limitation is the comparison with the PSS only. Associations with additional instruments for negative affect and chronic stress assessment should be implemented in future studies to fully examine convergent and divergent validity. Future research should demonstrate criterion validity by exploring the relationship with external ratings of chronic stress. A longitudinal design or a study with repeated measurements would provide opportunities for assessing the factor structures over time, determining retest-reliability, and testing for potential cohort effects.

Conclusions

In conclusion, our study presents a brief 9-item version of the TICS for measuring chronic stress that is theoretically based and shows strong reliability and emerging evidence for validity. Additionally, it features measurement invariance across participant gender allowing for meaningful interpretations of age and gender differences. Therefore, it can be recommended as an economical screening instrument for multivariable studies in psychology, medicine, and epidemiology.

Abbreviations

JAS = Job Anxiety Scale

WPS = Workplace Phobia Scale

CFI = Comparative Fit Index

TLI = Tucker-Lewis Index

RMSEA = Root Mean Square Error of Approximation

SRMR = Standardized Root Mean Square Residual

GH = Gamma Hat

ICD = International Statistical Classification of Diseases and Related Health Problems

DSM = Diagnostic and Statistical Manual of Mental Disorders

STAI-T = Stait-Trait-Anxiety Inventory

HEALTH = Hamburg Modules for the Assessment of Psychoisoical Health

Declarations

Ethics approval and consent to participate:

All participants volunteered and received a data protection declaration in agreement with the Helsinki Declaration. The gave both, written and verbal, informed consent. The study was approved according to the ethical guidelines by the Ethical commission of the Medical Faculty of the Technische Universität Dresden (EK 79032011).

Consent for publication:

Not applicable

Availability of data and material:

Data and materials are available from the corresponding author upon 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 TR supervised the process of creating this paper and with CB collected and provided data. EB and MS contributed substantially to conception and design. All authors have made substantial contributions to analysis and interpretation of data. BS and AH executed the statistical analyses. BS, KP, CB and TR drafted the manuscript and all authors revised it critically for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements:

Not applicable

References

1. Lohmann-Haislah A. Stressreport Deutschland 2012 - Psychische Anforderungen, Ressourcen und Befinden. Dortmund; 2012. Available from: <https://www.baua.de/DE/Angebote/Publikationen/Berichte/Gd68.html>
2. Becker P, Schulz P, Schlotz W. Persönlichkeit, chronischer Stress und körperliche Gesundheit. *Zeitschrift für Gesundheitspsychologie*. 2004;12(1):11–23.
3. Cohen S, Kessler RC, Gordon LU. *Measuring stress: a guide for health and social scientists*. New York: Oxford University Press; 1995.
4. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. *J Am Med Assoc*. 2007; 298(14):1685–7.
5. Schulz P, Hellhammer J, Schlotz W. Arbeitsstress, sozialer Stress und Schlafqualität: Differentielle Effekte unter Berücksichtigung von Alter, Besorgnisneigung und Gesundheit. *Zeitschrift für Gesundheitspsychologie*. 2003;11(1):1–9.
6. Chandola T, Brunner E, Marmot M. Chronic stress at work and the metabolic syndrome: Prospective study. *BMJ*. 2006;332(7540):521–4.
7. Roohafza H, Talaei M, Sadeghi M, Mackie M, Sarafzadegan N. Association Between Acute and Chronic Life Events on Acute Coronary Syndrome. *J Cardiovasc Nurs*. 2010;25(5):E1–7.
8. Ehrström S, Kornfeld D, Rylander E, Bohm-Starke N. Chronic stress in women with localised provoked vulvodinia. *J Psychosom Obstet Gynecol*. 2009;30(1):73–9.
9. Vedhara K, Irwin MR. *Human Psychoneuroimmunology*. New York: Oxford University Press; 2005.
10. Jackson JS, Knight KM, Rafferty JA. Race and Unhealthy Behaviors: Chronic Stress, the HPA Axis, and Physical and Mental Health Disparities Over the Life Course. *Am J Public Health*. 2010;100(5):933–9.

11. Levenstein S, Prantera C, Varvo V, Scribano ML, Berto E, Luzi C, et al. Development of the perceived stress questionnaire: A new tool for psychosomatic research. *J Psychosom Res.* 1993;37(1):19–32.
12. Fliege H, Rose M, Arck P, Walter OB, Kocalevent RD, Weber C, et al. The Perceived Stress Questionnaire (PSQ) reconsidered: Validation and reference values from different clinical and healthy adult samples. *Psychosom Med.* 2005;67(1):78–88.
13. Schulz P, Schlotz W, Becker P. Trierer Inventar zum chronischen Stress (TICS). Göttingen: Hogrefe; 2004. Available from: <https://www.testzentrale.de/shop/trierer-inventar-zum-chronischen-stress.html>
14. Petrowski K, Paul S, Albani C, Brähler E. Factor structure and psychometric properties of the trier inventory for chronic stress (TICS) in a representative German sample. *BMC Med Res Methodol.* 2012;12(42).
15. Becker P. Prävention und Gesundheitsförderung. In: Schwarzer R, editor. *Gesundheitspsychologie: ein Lehrbuch.* 2nd ed. Göttingen: Hogrefe-Verlag; 1997. p. 517–34.
16. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation; 1996.
17. Petrowski K, Kliem S, Albani C, Hinz A, Brähler E. Norm values and psychometric properties of the short version of the Trier Inventory for Chronic Stress (TICS) in a representative German sample. *PLoS One.* 2019;14(11).
18. 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(18).
19. Amazon Mechanical Turk [Internet]. [cited 2020 Sep 21]. Available from: <https://www.mturk.com/>
20. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav.* 1983;24(4):385–96.
21. Choi SW, Gibbons LE, Crane PK. lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *J Stat Softw.* 2011;39(8):1–30. Available from: <https://cran.r-project.org/package=lordif>
22. Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y. semTools: Useful tools for structural equation modeling. 2019. Available from: <https://CRAN.R-project.org/package=semTools>
23. Kelley K. MBESS: The MBESS R Package. 2019. Available from: <https://cran.r-project.org/package=MBESS>
24. Rosseel Y. lavaan: An R package for Structural Equation Modeling. *J Stat Softw.* 2012;48(2):1–36.
25. Beauducel A, Herzberg PY. On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Struct Equ Model A Multidiscip J.* 2006;13(2):186–203.
26. Flora DB, Curran PJ. An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychol Methods.* 2004;9(4):466–91.

27. Li C-H. The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol Methods*. 2016;21(3):369–87.
28. 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(2):23–74.
29. Hu L-T, Bentler PM. Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification. *Psychol Methods*. 1998;3(4):424–53.
30. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model A Multidiscip J*. 1999;6(1):1–55.
31. 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(3):399–412.
32. McDonald RP. *Test theory: A unified approach*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.; 1999. 485 p.
33. Milfont TL, Fischer R. Testing measurement invariance across groups: applications in cross-cultural research. *Int J Psychol Res*. 2010;3(1):111–30.
34. Wu H, Estabrook R. Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. *Psychometrika*. 2016;81(4):1014–45.
35. Kleinman M, Teresi JA. Differential item functioning magnitude and impact measures from item response theory models. *Psychol Test Assess Model*. 2016;58(1):79–98.
36. Raju NS, Fortmann-Johnson KA, Kim W, Morris SB, Nering ML, Oshima T. The Item Parameter Replication Method for Detecting Differential Functioning in the Polytomous DFIT Framework. *Appl Psychol Meas* . 2009;33(2):133–47.
37. Raju NS, Van der Linden WJ, Fler PF. IRT-Based Internal Measures of Differential Functioning of Items and Tests. *Appl Psychol Meas*. 1995;19(4):353–68.
38. Raju NS. *DFITP5: A Fortran Program for calculating dichotomous DIF/DTF*. Chicago: Illinois Institute of Technology; 1999.
39. Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. *Psychol Assess*. 2000;12(1):102–11.
40. Recklitis CJ, Yap L, Noam GG. Development of a Short Form of the Adolescent Version of the Defense Mechanisms Inventory. *J Pers Assess*. 1995;64(2):360–70.
41. Whitley RJ, Hromadka T V. Evaluating Uncertainty in Design Storm Runoff Predictions. *Water Resour Res*. 1991;27(10):2779–84.
42. Haynes SN, Richard DCS, Kubany ES. Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychol Assess*. 1995;7(3):238–47.

Figures

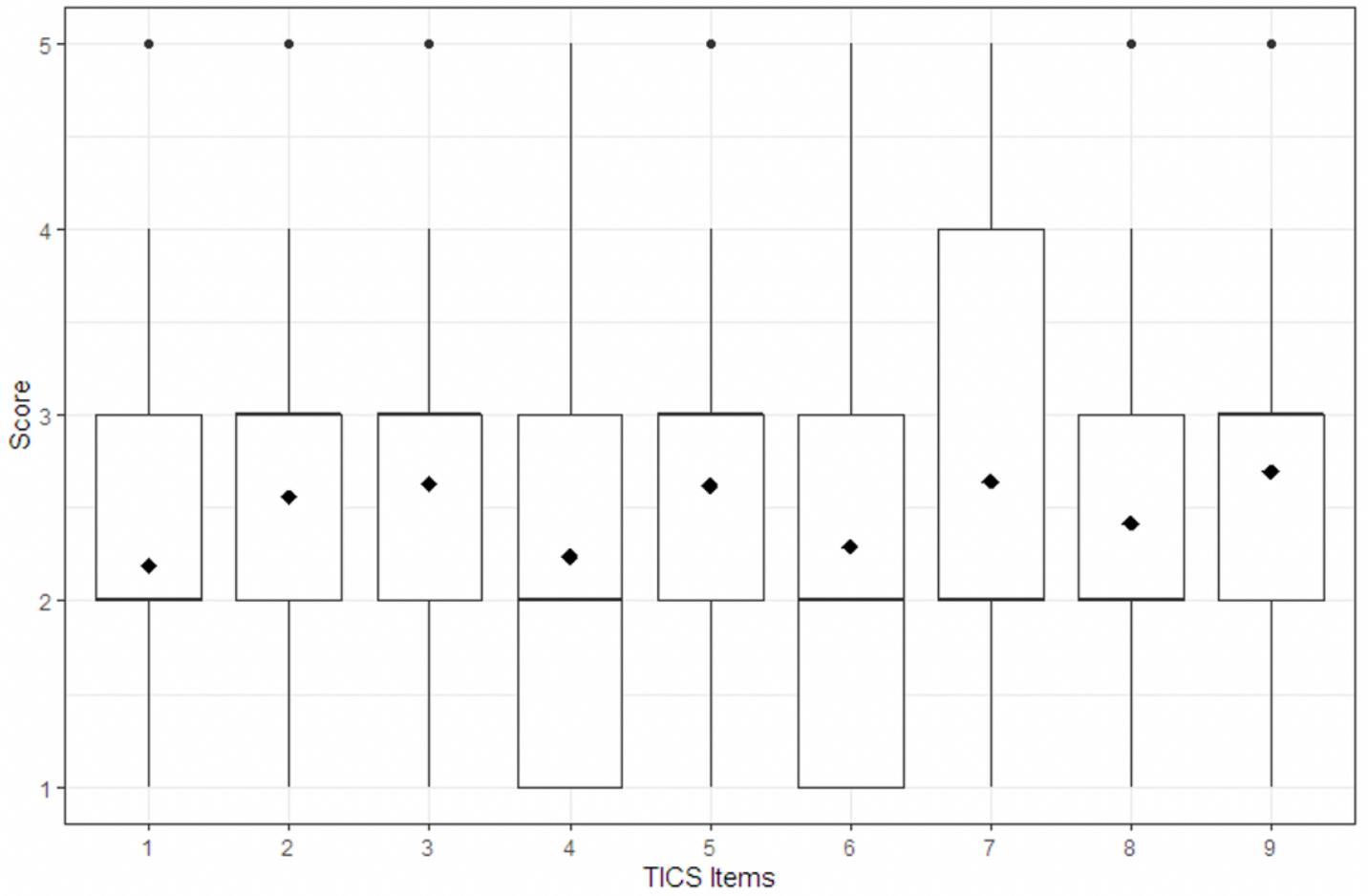


Figure 1

Boxplots of the TICS-9 item distributions. Diamonds represent the item score means.

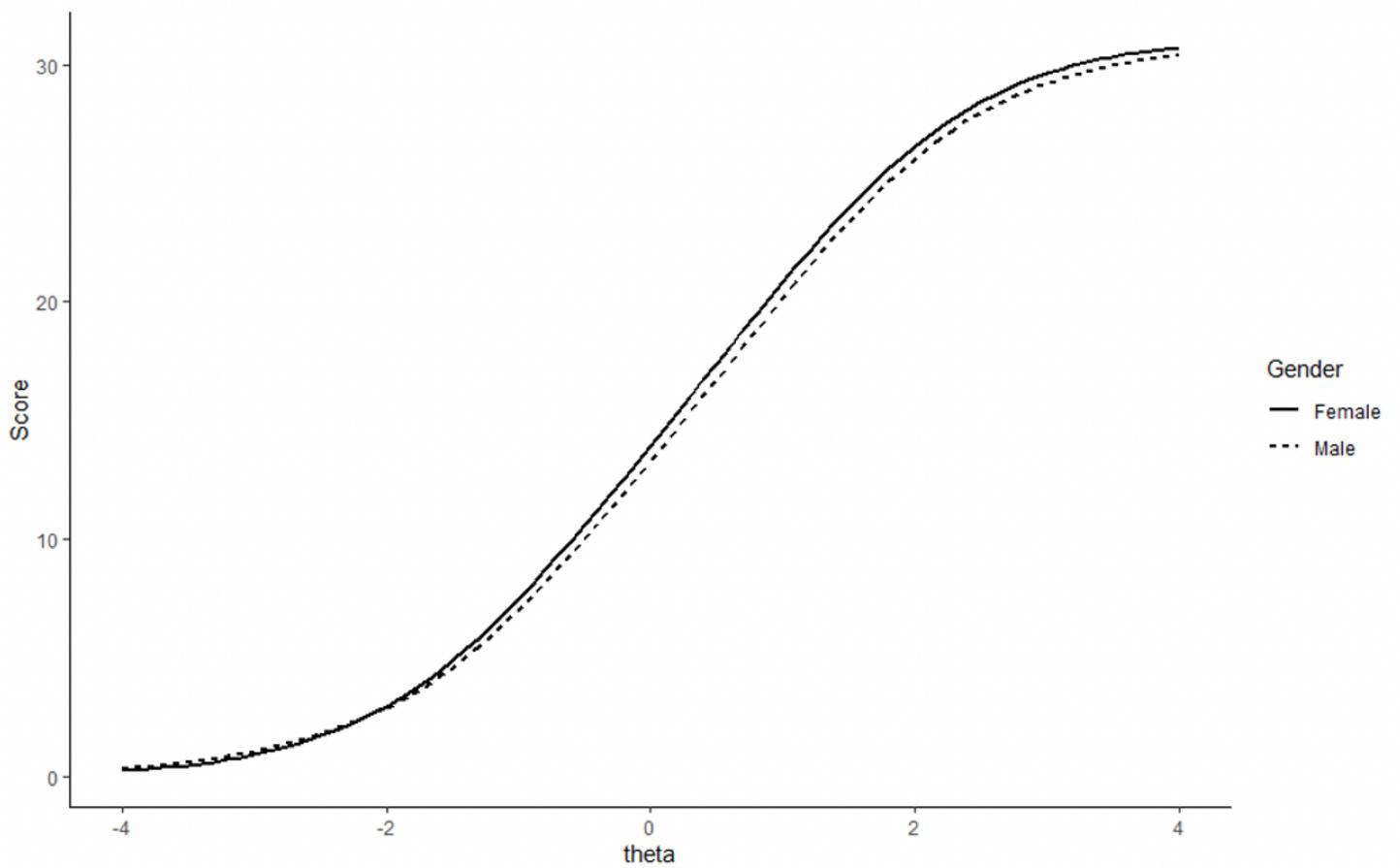


Figure 2

Gender-specific test characteristic curves. The scale range differs from the empirical distribution because items were rescaled to minima of 0 and groups of insufficient size were collapsed.