The Validity and Reliability of the Turkish Version of the Brief Irritability Test

ABSTRACT

Objective: Although irritability is a widely used term, it has no universal definition. Irritability is an emotional process that can be defined by a tendency to negative emotional states. No Turkish scale has been developed or adapted to measure irritability in adults. Consequently, this paper aims to conduct a validity and reliability study of the Turkish version of the Brief Irritability Test (BITe) with 5 items, which was developed by Holtzman et al. in 2015 to measure irritability rapidly and appropriately.

Methods: The Turkish BITe’s internal consistency and validity analysis were studied on 136 volunteering undergraduate and postgraduate students. Cronbach’s alpha value was calculated for internal consistency. Concurrent, convergent, discriminant validity analyses, and confirmatory factor analysis (CFA) were conducted to calculate structural validity. Moreover, the scale was applied to 24 people 2 weeks later to determine the temporal reliability of the Turkish BITe.

Results: When the fit indices of the scale related to the CFA were examined, it was observed that it had a good fit ($\chi^2 = 7.517, \chi^2/df = 1.503, df = 5; RMSEA = 0.061; CFI = 0.992, GFI = 0.977, NFI = 0.976; TLI = 0.984, IFI = 0.992$). In the reliability analysis, the Cronbach’s alpha value was 0.86, and the correlation coefficient between test–retest scores was 0.74 ($P < .001$).

Conclusion: This study reveals that the Turkish form of the BITe shows sufficient psychometric properties in the non-clinical population.

Keywords: Irritability, mood, depression, anxiety, stress

Introduction

The term “irritability” has no commonly agreed definition but is generally translated as nervousness, irascibility, or short temper, etc., Snaith and Taylor are one of the first to define irritability and they define it as “…Irritability is a feeling state characterised by reduced control over temper which usually results in irascible verbal or behavioural outbursts, although the mood may be present without observed manifestation. It may be experienced as brief episodes, in particular circumstances, or it may be prolonged and generalised. The experience of irritability is always unpleasant for the individual and overt manifestation lacks the cathartic effect of justified outbursts of anger.” It is not certain whether the word “irritability” matches the concept of “irritability” in clinical terms. In everyday practice the word “irritable” is being used to define a variety of conditions from psychiatric symptoms to patient behaviors. For example “irritable” may be the choice for describing patients who are simply angry and behaving in an aggressive manner and also patients who are delirious and violent. Malhi et al. evaluated 3 reasons for this difficulty in definition: (i) Everyone has their unique irritability experiences; (ii) The misuse of the term “irritability”; (iii) The combination of emotions, such as irritating agitation, anger, anxiety, confusion, and disappointment, leading to the difficulty in defining the term. Maybe as the result of this difficulty in describing the term “irritability,” it is being used interchangeably for defining other emotions. Irritability has emotional (defined as anger) and behavioral components (defined as aggression). With all these, the most up to date definition of irritability was made by...
Barata et al\(^4\) in 2016 and they defined it as “Irritability is an emotional process that is characterized by a proneness to experience negative affective states, such as anger, annoyance, and frustration, which may or may not be outwardly expressed. Irritability often includes a feeling that one’s emotional responses are unjustified or disproportionate to the immediate source, but difficult to control.”

Irritability is a condition seen in many psychiatric and non-psychiatric medical conditions.\(^5\) When evaluating the relationship between irritability and psychiatric disorders, it is important to consider gender, age, personality structure, social relations, environmental conditions, brain pathologies, and pharmacological and endocrinological factors.\(^1\) Although irritability has a place in the diagnostic criteria for psychiatric disorders, brain pathologies, and pharmacological and endocrinological processes, it is important to consider gender, age, personality structure, social relations, environmental conditions, brain pathologies, and pharmacological and endocrinological factors.\(^1\) Although irritability has a place in the diagnostic criteria for major depressive disorder in children and adolescents in DSM-5, the same is not valid for adults.\(^6\) Meanwhile, 40% of patients with non-psychotic major depressive episodes have significant irritability.\(^7\) Irritability is also common in anxiety disorders, such as generalized anxiety disorder, obsessive-compulsive disorder, panic disorder, post-traumatic stress disorder, and social phobia. High levels of irritability are common in females and younger people with previous frequent suicidal attempts and thoughts. This group has been reported to have a lower quality of life and experience more depression and anxiety.\(^8\) Stringaris et al\(^9\) also found that irritability detected in early adolescence predicts major depressive disorder, generalized anxiety disorder, and dysthymia in adulthood but not bipolar disorder and personality disorders.

In the evaluation of irritability, which is a significant phenomenon, measurement methods based on a single question, such as the Beck Depression Inventory and the items of the Young Mania Rating Scale, are generally used.\(^10,11\) Scales measuring irritability with more than one question have also been developed.\(^12,13\) These relatively long scales provide strong reliability and evaluate other structures other than irritability, including aggression, hostility, and depression.\(^14\) The DSM-5 Level 2 irritability scale for children and adults was adapted to Turkish in 2017 by Yalın Sapmaz et al\(^15\) in Turkey. As far as we know, in the literature, Turkey has no scale developed or adapted to measure irritability in adults. It has been shown that the 5-item Likert-type scale, developed by Holtzman et al\(^16\) in 2015, measures irritability briefly and appropriately.

Hence, this paper aims to conduct a Turkish validity and reliability study of the Brief Irritability Test (BITe), which was shown to measure irritability briefly and reliably. This paper hypothesizes that the Turkish version of the BITe will be valid and reliable.
Quality of Life Short Form-36 (SF-36): This is a 36-item self-report scale consisting of physical function, social function, role limitations, mental health, vitality, pain, and general perception of health.\(^{24}\) Higher scores on all subscales in this scale indicate better health. Koçyiğit et al\(^{25}\) conducted its Turkish validity and reliability study in 1999.\(^{25}\)

Statistical Analysis
SPSS AMOS 23 version (IBM Corp., Armonk, NY, USA) was used for the confirmatory factor analysis (CFA) process to be applied to test the construct validity of the Turkish BITe. Standard estimation method was used for the CFA. The validity of the models is evaluated with the goodness of fit tests of the data.\(^{26}\) The fit indices used in the research are relative chi-square index (chi-square fit index/degrees of freedom \(\chi^2/df\)), comparative fit index (CFI),\(^{27}\) general fit index (GFI), normalized fit index (NFI), and root mean square error (RMSEA).\(^{28}\) The CFI, GFI, NFI \(>0.900, \chi^2/df < 5\), and RMSEA \(<0.0854\) values can be used as acceptable good-fit criteria.\(^{29}\)

For other validity and reliability analyses in the study, SPSS version 22.0 (IBM Corp., Armonk, NY, USA) was used. Descriptive statistics of the data are presented with \(n\) (%) and, for non-normalized variables are shown as “median (min-max),” and normal distributions are shown as mean (SD). The conformity of the data to the normal distribution was examined with Shapiro–Wilks’s test. Cronbach’s alpha coefficient was calculated and it was found to be 0.86. The scale was re-applied to 24 participants 2 weeks later to determine the temporal reliability of the BITe. The test–retest correlation structure was examined with an exploratory factor analysis (EFA) process to be applied to test the construct validity of the Turkish BITe. Standardized regression coefficients calculated as a result of EFA are shown in Figure 2. When the fit indices for the analysis were examined, it was seen that they fit well. Moreover, all factor loadings were significant for the items (\(P < .001\)). Table 2 shows the CFA findings of the scale.

Results
Descriptive Statistics
The sample consists of 136 participants, 41 males (30.1%) and 95 females (69.9%). The mean age of the sample is 23.3 (SD = 3.11). Of the participants, 128 were single (94.1%) and 8 were married (5.9%).

Reliability Analyses
Cronbach’s alpha coefficient was calculated to determine the internal consistency level, test–retest correlation showing temporal invariance, and item-total score analysis was used to test the reliability of the Turkish version of the BITe. The relationships between the scales were analyzed using the Pearson correlation coefficient to conduct the concurrent, convergent, and discriminant validity analyses.

Concurrent, Convergent, and Discriminant Validity Analyses
For validity analyses, concurrent, convergent, and discriminant validity analyses were performed after the CFA. Table 3 shows the correlation of the Turkish BITe with other applied scales. In the concurrency analysis, the BITe is expected to be related to depression, anxiety, stress, and other irritability scales. As seen in Table 3, the Turkish version of the BITe was statistically significantly correlated with the DASS-21 depression (\(r = 0.51; P < .001\)), anxiety (\(r = 0.42; P < .001\)), and stress (\(r = 0.71; P < .001\)) subscales, the BPAQ’s hostility (\(r = 0.41; P < .001\)), anger (\(r = 0.57; P < .001\)), verbal aggression (\(r = 0.30; P < .001\)), and physical aggression (\(r = 0.17; P < .037\)) and the trait anger subscale of STAXI (\(r = 0.55; P < .001\)).

Structure Validity Analyses
The Keiser–Meyer–Olkin (KMO) sampling adequacy statistic was calculated as KMO = 0.83, and the sample size was found to be sufficient. According to the results of the Bartlett sphericity test, it was determined that there was a sufficient level of correlation between the items for exploratory factor analysis (\(\chi^2 = 228; P < .001\)). The factor structure was examined with an exploratory factor analysis using a condition of Eigenvalues bigger than 1 rule. It was observed that BITe could explain 63.63% of the total variance in a single subdimension. There was one-factor with an eigenvalue greater than 1 (Eigenvalue = 3.181). One-factor construct can also be seen in Figure 1.

CFA was conducted to test the single-factor model of the original scale. Standardized regression coefficients calculated as a result of CFA are shown in Figure 2. When the fit indices for the analysis were examined, it was seen that they fit well. Moreover, all factor loadings were significant for the items (\(P < .001\)). Table 2 shows the CFA findings of the scale.

Table 1. The Results of the Reliability Analysis of the Turkish Form of the BITe

| Items                                      | Mean (SD) | Corrected Item-Total Correlation | Cronbach’s Alpha If Item Deleted |
|--------------------------------------------|-----------|----------------------------------|---------------------------------|
| 1. I have been grumpy                       | 2.55 (0.89) | 0.64                              | 0.82                            |
| 2. I have been feeling like I might snap    | 2.23 (0.97) | 0.68                              | 0.83                            |
| 3. Other people have been getting on my nerves | 2.85 (0.93) | 0.70                              | 0.82                            |
| 4. Things have been bothering me more than they normally do | 2.82 (1.15) | 0.56                              | 0.86                            |
| 5. I have been feeling irritable            | 2.54 (1.06) | 0.80                              | 0.79                            |
In the discriminant validity analysis, it can be expected that social support and quality of life have a negative relationship with the BITe. For this, the relationship between the MSPSS and SF-36 and the BITe was examined. As displayed in Table 3, a statistically significant negative correlation was found with MSPSS total score ($r = -0.18; P = 0.028$), role physical ($r = -0.36; P < .001$), bodily pain ($r = -0.23; P = 0.006$), general health ($r = -0.45; P < .001$), vitality ($r = -0.53; P < .001$), social functioning ($r = -0.39; P < .001$), role emotional ($r = -0.22; P = 0.008$), and mental health ($r = -0.65; P < .001$).

**Discussion**

This study aimed to carry out the validity and reliability study of the Turkish version of the BITe, which measures irritability briefly and reliably. First, CFA was applied to test the construct validity of the Turkish BITe. The single-factor structure was used in the CFA as in the original article. According to the fit indices obtained, the fit indices of the single-factor model of the BITe was calculated as $\chi^2/df = 1.503$, $df = 5$; RMSEA = 0.061; CFI = 0.992, GFI = 0.977, NFI = 0.976; Tucker–Lewis Index (TLI) = 0.984, incremental fit index (IFI) = 0.992. In the original article, only the GFI fit index was calculated, and it was 0.99. Considering the goodness of fit index values, the structure validity of the 5-item single-factor structure of the Turkish version of the BITe is sufficient.

The reliability analyses of the scale were conducted after the construct validity analysis. The calculated Cronbach’s alpha value of the scale was 0.86 and omega value was 0.86. In the article on the development of the scale, Cronbach’s alpha value was 0.88 and omega value was not calculated. While most publications find a Cronbach’s alpha value higher than 0.70 acceptable, there are publications stating that the Cronbach’s alpha value should be 0.80 or higher for scales of 3-10 items. From this perspective, the BITe has good internal consistency. When the test–retest correlation between the total scores of the scale was examined, the correlation coefficient was calculated as 0.74 at a moderate level. The test–retest correlation was not calculated in the original study, one of its limitations. When evaluated in its current form, it can be said that the Turkish version of the BITe has sufficient temporal consistency.

For the concurrent validity analysis, the BITe’s relationship with depression, anxiety, stress, and other irritability scales was examined. Accordingly, a statistically significant low-to-moderate correlation was found between the scales examined and the BITe ($r = 0.17$-$0.71$).

Table 2. Confirmatory Factor Analysis Findings of the Turkish Form of the BITe

| $K$ | $\chi^2$ | $df$ | $\chi^2/df$ | $P$ | CFI | GFI | NFI | RMSEA | SRMR |
|-----|---------|------|-------------|-----|-----|-----|-----|-------|------|
| 5   | 7.517   | 5    | 1.503       | <.001 | 0.992 | 0.977 | 0.976 | 0.061 | 0.0310 |

$k$, number of items; $df$, degree of freedom; CFI, comparative fit index; GFI, goodness of fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean residual; NFI, normed fit index.

Table 3. Pearson Correlation Coefficients Among the Turkish Form of the BITe and Other Scales

| Scales (n = 136) | Mean (SD) | Pearson Correlation Coefficients | $P$  |
|-----------------|-----------|----------------------------------|------|
| DASS-21 Depression | 5.46 (4.04) | 0.51 <.001 |      |
| DASS-21 Stress | 5.46 (3.43) | 0.71 <.001 |      |
| DASS-21 Anxiety | 3.22 (2.89) | 0.42 <.001 |      |
| BPAQ Physical Aggression | 8.59 (6.08) | 0.17 .037 |      |
| BPAQ Hostility | 12.73 (5.24) | 0.41 <.001 |      |
| BPAQ Anger | 10.25 (5.41) | 0.57 <.001 |      |
| BPAQ Verbal Aggression | 8.91 (3.46) | 0.30 <.001 |      |
| STAXI T-Anger | 17.83 (4.38) | 0.55 <.001 |      |
| MSPSS T-Score | 61.59 (13.60) | 0.26 .002 |      |
| SF-36 Physical Functioning | 93.67 (9.87) | 0.18 .028 |      |
| SF-36 Role Physical | 79.65 (35.51) | 0.36 <.001 |      |
| SF-36 Bodily Pain | 70.58 (17.03) | 0.23 .006 |      |
| SF-36 General Health | 66.86 (19.28) | 0.45 <.001 |      |
| SF-36 Vitality | 51.98 (21.88) | 0.53 <.001 |      |
| SF-36 Social Functioning | 71.23 (23.28) | 0.39 <.001 |      |
| SF-36 Role Emotional | 50.98 (42.15) | 0.22 .008 |      |
| SF-36 Mental Health | 62.82 (17.26) | 0.65 <.001 |      |

DASS-21, Depression Anxiety Stress Scale-21; BPAQ, Buss Perry Aggression Questionnaire; MSPSS, Multidimensional Scale of Perceived Social Support; STAXI, State-Trait Anger Expression Inventory; SF-36, Quality of Life Short Form-36.
In the development article of the scale, it was reported that BITe correlated with anger, depression, aggression, quality of life, and neuroticism scales. The lowest correlations found were 0.18 on the BPAQ physical aggression subscale and 0.30 on the BPAQ verbal aggression subscale. The highest correlation was with the DASS-21 stress subscale. Some authors, such as Snaith and Taylor, consider irritability as a mood rather than an attitude and behavior. For the discriminant validity analysis, the correlation of BITe with social support and quality of life was examined. In the scales examined, a statistically significant correlation between low and moderate was found among BITe (r = 0.18-0.65). The higher correlation of the BITe with scales measuring moods, such as depression and anger, rather than scales measuring tendency to behavior, such as hostility and aggression, can be interpreted as the evaluation of irritability as a mood or as BITe’s measuring the emotional dimension of irritability.

Some scales used for convergent/discriminant validity analysis were similar with the original study. BPAQ’s subscales anger, verbal aggression, physical aggression, and hostility were also evaluated in the original study and their Pearson correlations with BITe were 0.56, 0.25, 0.25, and 0.52, respectively. These results are similar as well. Also, it is important to note that lowest correlations in BPAQ were physical aggression and verbal aggression in both studies and were statistically significant.

This study has several limitations. The first is that the study was conducted in a non-clinical population. Additional studies are needed for the validity and reliability of the BITe in the clinical population. Second, 69.9% of this study’s sample were women. This may limit the generalization of results to men. Third, a low test–retest sample size may be considered as a limitation of the study. Moreover, the fact that the study was conducted online may create a limitation in terms of data reliability.

The findings of this study to show the Turkish validity and reliability of the BITe, which was developed to measure irritability briefly and reliably, show that the BITe has sufficient psychometric properties in the non-clinical population.

References
1. Snaith RP, Taylor CM. Irritability definition, assessment and associated factors. Br J Psychiatry. 1985;147:127-136. [CrossRef]
2. Malhi G, Bell E, Outhred T. Getting irritable about irritability? Evid Based Ment Health. 2019;22(3):93-94. [CrossRef]
3. Caprara GV, Cinanni V, D’imperio G, et al. Indicators of impulsive aggression: present status of research on irritability and emotional susceptibility scales. Pers Individ Dif. 1985;6(6):665-674. [CrossRef]
4. Barata PC, Holtzman S, Cunningham S, O’Connor BP, Stewart DE. Building a definition of irritability from academic definitions and lay descriptions. Emot Rev. 2016;8(2):164-172. [CrossRef]
5. Mangelli L, Fava GA, Grassi L, et al. Irritable mood in Italian patients with medical disease. J Nerv Ment Dis. 2006;194(3):226-228. [CrossRef]
6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Virginia: American Psychiatric Association; 2013. [CrossRef]
7. Perlis RH, Fraguas R, Fava M, et al. Prevalence and clinical correlates of irritability in major depressive disorder: a preliminary report from the sequenced treatment alternatives to relieve depression study. J Clin Psychiatry. 2005;66(2):159-66; quiz 147, 273. [CrossRef]
8. Perlis RH, Fava M, Trivedi MH, et al. Irritability is associated with anxiety and greater severity, but not bipolar spectrum features, in major depressive disorder: Acta Psychiatr Scand. 2009;119(4):282-289. [CrossRef]
9. Stringaris A, Cohen P, Pine DS, Leibenluft E. Adult outcomes of youth irritability: a 20-year prospective community-based study. Am J Psychiatry. 2009;166(9):1048-1054. [CrossRef]
10. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.
11. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry. 1978;133(11):429-435. [CrossRef]
12. Craig KJ, Hietanen J, Markova IS, Berrios GE. The irritability questionnaire: a new scale for the measurement of irritability. Psychiatry Res. 2008;159(3):367-375. [CrossRef]
13. Stringaris A, Goodman R, Ferdinando S, et al. The affective reactivity index: a concise irritability scale for clinical and research settings. J Child Psychol Psychiatry. 2012;53(11):1109-1117. [CrossRef]
14. Holtzman S, O’Connor BP, Barata PC, Stewart DE. The brief irritability test (BITe): a measure of irritability for use among men and women. Assessment. 2015;22(1):101-115. [CrossRef]
15. Yalın Sapmaz Ş, Yalın N, Ergin C, et al. Validity and reliability of DSM-5 level 2 irritability scale-Turkish version. Article in Turkish. Çocuk Genceğir Ruh Sağlığı Derg. 2017;24(1):15-23.
16. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol. 2005;44(2):227-239. [CrossRef]
17. Yılmaz Ö, Boz H, Arslan A. The validity and reliability of Depression Stress and Anxiety Scale (DASS21) Turkish short form. Article in Turkish. J Financ Econ Soc Res. 2017;2(2):78-91.
18. Buss AH, Perry M. The aggression questionnaire. J Pers Soc Psychol. 1992;63(3):452-459. [CrossRef]
19. Demirtaş Madran HA. Reliability and validity of the Buss-Perry Aggression Questionnaire-Turkish version. Article in Turkish. Türk Psikiyatr Derg. 2013;24(2):124-129.
20. Zimet GD, Dahlem NW, Zimet SG, Farley GK. The multidimensional scale of perceived social support. J Pers Assess. 1988;52(1):30-41. [CrossRef]
21. Eker D, Arkar H, Yaldız H. Factorial structure, validity, and reliability of revised form of the multidimensional scale of perceived social support. Article in Turkish. Türk Psikiyatr Derg. 2001;12:17-25.
22. Spielberger CD. State-Trait Anger Expression Inventory [Published online]. Odessa, FL: Psychological Assessment Resources Inc.; 1988.
23. Özer AK. Sürekli öfke (SL-ÖFKE) ve öfke ifade tarzı (ÖFKE-TARZ) ölçekleri ön çalışması. Türk Psikol Derg. 1994;9(31):26-35.

24. Ware Jr JE, Sherbourne CD, The Maryland Ornithological Society. 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med Care. 1992;3(6):473-483.

25. Koçyiğit H, Aydemir O, Ölmez N, Memiş A. Kısa form-36 (KF36)'ın Türkçe versiyonunun güvenilirliği ve geçerliliği [Reliability and validity of the Turkish version of the short form-36]. Article in Turkish. İlaç ve Tedavi Derg. 1999;12(2):102-106.

26. Brown TA. Confirmatory Factor Analysis for Applied Research. New York: Guilford Publications; 2015.

27. Bentler PM. Comparative fit indexes in structural models. Psychol Bull. 1990;107(2):238-246. [CrossRef]

28. Hu LT, Bentler PM. Evaluation model fit. In: Hoyle RH, ed. Structural Equation Modeling: Concepts, Issues and Application. Thousand Oaks, CA: Sage; 1995:77-99.

29. Şimşek OF. Yapısal Eşitlik Modellemesine Giriş: (Temel İlkeler ve LISREL Uygulamaları) Book in Turkish. İstanbul: Ekinoks; 2007.

30. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol. 1993;78(1):98-104. [CrossRef]

31. Pallant J. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS. London: Routledge; 2020.