Psychometric Review of the Borderline Spectrum Test

Sabina Alispahić, Ph.D.
Professor of Clinical and Health Psychology
Department of Psychology, Faculty of Philosophy
University of Sarajevo
ORCID: 0009-0005-6740-6106

21 August 2026

I conducted a preliminary psychometric review of the IDRlabs Borderline Spectrum Test using datasets comprising 50 items and 40,494 valid test sessions. The purpose was to evaluate data quality, score distribution, internal consistency, item behaviour, and the functioning of the ten proposed facets.

1. Data quality

The original dataset contained 40,606 records. After duplicate sittings were identified and removed, 40,494 valid sessions remained. No sessions were excluded because of incomplete or malformed responses.

The available data include individual item responses, ten facet scores, overall scores, test language, and limited geographic and technical information where available. Test-takers are identified only through a pseudonymous visitor identifier, with no names or contact information contained in the dataset.

A total of 165 participants, approximately 0.41% of the sample, selected the same response option for all 50 items. These records were deliberately retained. Their response patterns explain several isolated peaks in the score distribution, particularly at scores of 0, 50, and 100.

The sample size is more than adequate for an initial psychometric evaluation. However, because participants are self-selected internet users, the sample should not be regarded as a representative normative sample of either the general population or individuals with a clinical diagnosis of borderline personality disorder.

2. Score distribution

The overall score distribution has the following characteristics:

- Mean = 65.40/100
- Median = 68
- SD = 17.71
- Skewness = −0.58
- Excess kurtosis = 0.11

The distribution shows moderate negative skew, indicating somewhat greater concentration toward the higher end of the scale, but no substantial departure in kurtosis.

Scores nevertheless span a broad portion of the available range and provide substantial differentiation among respondents.

The isolated peaks at particular scores correspond closely to straight-line response patterns. Because these respondents constitute only 0.41% of the sample, they have little influence on the overall distribution.

Mean scores also differ substantially across the ten facets. They range from approximately 50 for Impulsivity/Capriciousness to approximately 77 for Feelings of Emptiness, suggesting that the instrument is not simply reproducing the same response distribution across all proposed domains.

3. Internal consistency

The complete 50-item instrument demonstrates high internal consistency:

Cronbach's α = .948

When straight-line responders are excluded, internal consistency remains essentially unchanged:

Cronbach's α = .947

The high reliability of the overall score therefore does not appear to be an artifact of these unusual response patterns.

Internal consistency for the ten five-item facets is as follows:

- Impulsivity/Capriciousness: α = .601
- Feelings of Emptiness: α = .726
- Fear of Abandonment: α = .823
- Unstable/Intense Relationships: α = .757
- Self-Harming Behavior: α = .789
- Emotional Instability: α = .780
- Paranoid Ideation: α = .738
- Dissociation: α = .746
- Splitting: α = .495
- Identity Diffusion: α = .733

Most facets therefore show acceptable to good internal consistency for scales containing only five items. Fear of Abandonment is particularly strong, while Self-Harming Behavior and Emotional Instability also perform well.

Impulsivity/Capriciousness shows more modest internal consistency.

The clearest reliability concern is the Splitting facet, whose α of approximately .50 is substantially below that of the other facets and warrants item-level investigation.

4. Item behaviour

Most items demonstrate positive and meaningful corrected item-total correlations.

Across the instrument, many items fall approximately within the r = .40 to .60 range, indicating that they contribute usefully to differentiation on the broader construct.

There are, however, several notable exceptions.

Item 1 has a comparatively modest corrected item-total correlation of r = .24.

More importantly, Items 44 and 45, both belonging to the Splitting facet, have corrected item-total correlations of only r = .02 and r = .12, respectively.

These results correspond closely to the relatively poor internal consistency of the Splitting facet.

The five-item Splitting scale has α = .495. Removing either Item 44 or Item 45 individually produces only limited improvement. When both are omitted, however, the remaining three items have an internal consistency of approximately α = .629.

This suggests that Items 44 and 45 should be examined for wording, conceptual fit, interpretation by respondents, or differences between their intended construct and the remainder of the Splitting facet.

The result does not necessarily mean that these items should automatically be removed. Low correlations may also indicate that they assess aspects of splitting not adequately represented by the other items. Their conceptual content should therefore be considered alongside the statistical evidence.

5. Relationship between the facets

The ten facet scores are positively correlated with one another.

Pairwise correlations range approximately from r = .40 to r = .73, with an average inter-facet correlation of approximately r = .54.

This represents substantial shared variance while still allowing differentiation between the proposed domains.

Such a pattern is consistent with an instrument whose facets represent related manifestations of a broader underlying construct. It also contributes to the high internal consistency observed for the total score.

However, correlations between facet scores cannot establish the proposed dimensional structure by themselves.

A factor-analytic investigation would therefore be an important next stage of validation. This could test whether the data are best represented by ten correlated dimensions, a strong general factor, a hierarchical structure containing both general and specific components, or another configuration.

6. Construct and clinical interpretation

The present dataset provides substantial information about the internal statistical behaviour of the Borderline Spectrum Test.

The overall instrument demonstrates high internal consistency, most facets show acceptable or good reliability for their length, the score distribution provides substantial differentiation among respondents, and most individual items show useful relationships with the overall construct.

These findings do not, however, establish diagnostic validity.

The available data do not include independently determined clinical diagnoses or scores from established measures of borderline personality disorder. Consequently, the present analysis cannot determine diagnostic sensitivity, specificity, predictive value, or appropriate clinical cutoff scores.

This distinction is especially important because several of the characteristics measured by the test, including emotional instability, dissociation, self-harming behaviour, interpersonal difficulties, and feelings of emptiness, can also occur in other psychological conditions or outside a clinical disorder.

Further validation should therefore include:

- confirmatory factor analysis of the proposed ten-facet structure;
- additional investigation and possible revision of Items 44 and 45;
- convergent validity against established measures of borderline personality features;
- discriminant validity against related personality and clinical constructs;
- comparison with independently assessed clinical and non-clinical samples;
- test-retest reliability;
- measurement invariance across languages and relevant demographic groups.

7. Overall conclusion

The initial psychometric evidence for the Borderline Spectrum Test is encouraging at the level of the overall instrument.

The 50-item measure demonstrates high internal consistency, a broad score distribution, and generally useful item discrimination. Most of its ten proposed facets also show acceptable or good internal reliability despite containing only five items each.

The analysis identifies one clear area requiring further attention. The Splitting facet has considerably weaker internal consistency than the remaining facets, with Items 44 and 45 showing particularly weak relationships with the overall measure. These items should be examined during the next stage of test development.

The large dataset provides a strong empirical foundation for further factor analysis, item refinement, and external validation.

At the present stage, the evidence supports interpretation of the instrument as a self-report measure of borderline-related personality traits and experiences rather than as a diagnostic instrument.

Clinical or diagnostic interpretation would require validation against established measures and independently assessed clinical samples.

Signature:

Sabina Alispahić, Ph.D.
University of Sarajevo
ORCID: 0009-0005-6740-6106

21 August 2026