

Psychometric Review of the Racism 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

25 August 2026

I conducted a preliminary psychometric review of the IDRLabs Racism Test using the supplied results package covering 20 items, three proposed facets, and 37,477 valid test sessions. The purpose was to evaluate data quality, score distribution, internal consistency, item behaviour, and the extent to which the three proposed facets function as related components of the broader measure.

1. Data quality

The dataset is large for an initial psychometric evaluation. The original export contained 37,594 records. After same-visitor identical-answer duplicates were collapsed, 37,477 valid test sessions remained. The test was represented across 17 published locales.

The supplied package indicates that incomplete or out-of-range sittings were excluded and that straight-line response patterns were retained. There were 329 straight-line sittings, approximately 0.88% of the valid sample. This is low enough that it is unlikely to determine the aggregate results, although future validation work should still specify an a priori response-quality criterion.

The package supplied for this review contains aggregate psychometric outputs, item statistics, and generated figures rather than the session-level response matrix. The statistics below therefore describe and interpret the supplied outputs; raw-data sensitivity analyses cannot be independently recomputed from this package alone.

The sample size is more than adequate for psychometric analysis. However, participants are self-selected internet users rather than a probability sample, so these data should not be treated as representative population norms or as estimates of the prevalence of racist attitudes or behaviour.

2. Score distribution

The overall score distribution is statistically well behaved:

- Mean = 43.14/100
- Median = 42.5
- SD = 22.92
- Skewness = 0.25
- Excess kurtosis = -0.64

The distribution shows only modest positive skew and retains broad dispersion across the scale. The 5th and 95th percentiles are 9 and 84, respectively, with no evidence of severe overall floor or ceiling compression.

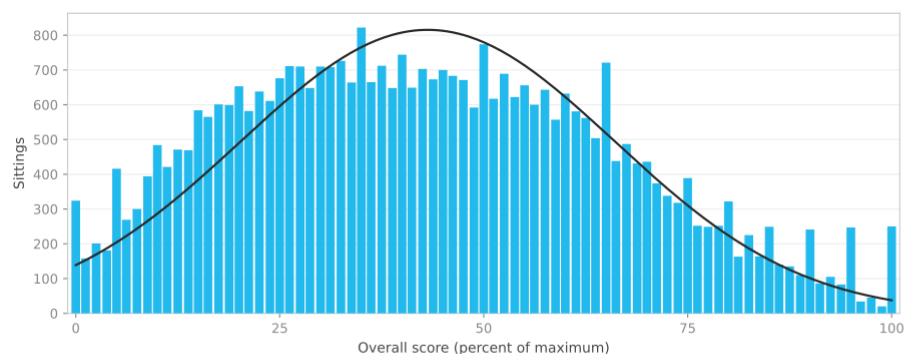


Figure 1. Overall-score distribution from the supplied results package.

The three facet means span a meaningful range, from 36.26 for Socio-Historical Racism to 55.01 for Perpetual Foreigner Racism. This indicates that the measure is not simply producing the same score level across all proposed domains.

3. Internal consistency and item behaviour

Internal consistency is strong at the level of the full 20-item measure.

Cronbach's alpha = .917

Internal consistency for the three facets is as follows:

- General Racism (6 items): alpha = .780
- Socio-Historical Racism (10 items): alpha = .868
- Perpetual Foreigner Racism (4 items): alpha = .695

The first two facet values are solid. Perpetual Foreigner Racism is slightly below the conventional .70 benchmark, but it contains only four items and the shortfall is small, so it should not be interpreted mechanically.

Corrected item-total correlations range from $r = .42$ to $.76$ across the 20 items, and all supplied first-component loadings are positive. There is therefore no obvious item that behaves in the wrong direction or shows extremely weak discrimination in the supplied results.

4. Facet structure

The three facet scores are substantially correlated. Pairwise correlations range from $r = .61$ to $.77$, with the strongest association between General Racism and Socio-Historical Racism.

This pattern is compatible with related but non-identical domains, while also suggesting a strong broader common dimension. The correlations are high enough that the distinct value of reporting all three subscales should be tested directly rather than assumed.

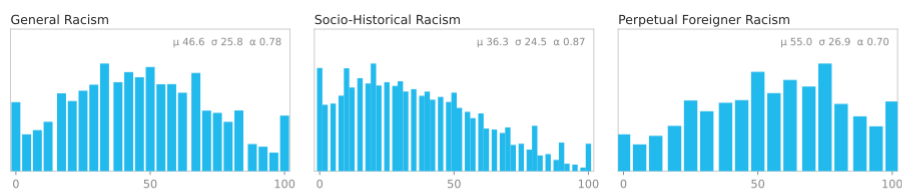


Figure 2. Facet-score distributions from the supplied results package.

Confirmatory factor analysis would be needed to determine whether the proposed three-facet model fits the item-level data better than a one-factor, higher-order, or bifactor representation.

5. Construct interpretation

The available results support the internal coherence of a self-report measure of attitudes and acceptability judgments related to racism. They do not establish that the score is a direct measure of discriminatory behaviour, a population prevalence estimate, or an externally validated classification.

The package also reports a 3.7-point difference between mobile and desktop mean scores. This does not by itself establish a measurement problem, but it is large enough to justify checking mode effects and measurement invariance in future work.

Further validation should ideally examine:

- convergent validity against established measures of racial attitudes and prejudice;
- discriminant validity against broader hostility, authoritarianism, social desirability, and related constructs;
- criterion validity using behavioural, informant, or experimentally derived indicators where feasible;
- confirmatory factor analysis comparing one-factor, three-factor, higher-order, and bifactor models;
- test-retest reliability;
- measurement invariance across language, sex, age, device type, and other relevant groups.

6. Overall conclusion

The initial psychometric evidence for the Racism Test is encouraging. The 20-item instrument demonstrates strong overall internal consistency, generally acceptable-to-good facet reliability, positive item discrimination throughout the supplied item set, and a broad total-score distribution suitable for further psychometric work.

The principal unresolved issues concern external and structural validity rather than obvious basic scoring problems. The three facets are fairly strongly intercorrelated, so their distinctness should be tested formally, and the scores should not be treated as population norms or direct measures of real-world discriminatory behaviour.

Subject to those limitations, the available evidence supports continued development and formal validation of the measure.

Signature:



Sabina Alispahić, Ph.D.

25 August 2026