

Psychometric Review of the Jealousy Test

Sabina Alispahić, Ph.D.
Professor of Clinical and Health Psychology, University of Sarajevo

ORCID: 0009-0005-6740-6106 • 29 September 2026

Overall assessment

The supplied results provide an encouraging basis for the [IDRlabs Jealousy Test](#) as an educational self-report measure of cognitive, emotional, and behavioral jealousy. Across 367,818 completed sittings, the 24-item test shows strong internal consistency, positive item discrimination throughout, and substantial variation in scores. Each of its three eight-item facets also has good internal consistency.

The appropriate assessment is optimistic. These findings support continued use for self-reflection and further validation, but do not establish diagnostic accuracy, representative population norms, or the distinctness of the three proposed dimensions.

Internal consistency

Overall Cronbach’s alpha is .943 to .94. Facet estimates range from .86 to .90. No facet shows an obvious internal-consistency weakness in the supplied outputs.

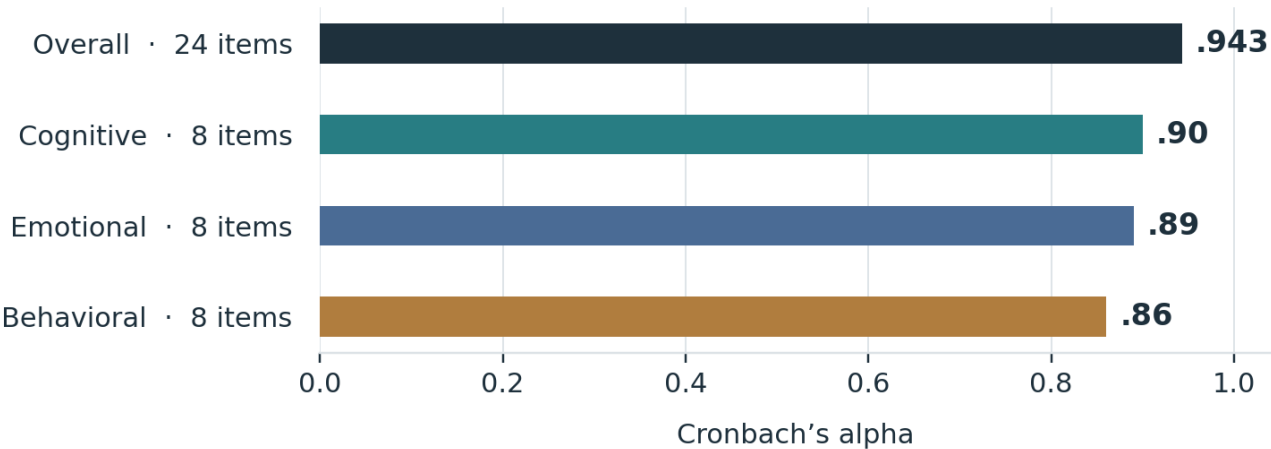


Figure 1. Reported internal consistency for the overall score and three facets. [1]

This is a clear strength of the current evidence. However, alpha describes how consistently items covary in this sample; it does not demonstrate a single underlying factor or validity of the score’s intended interpretation. Shared wording, response style, and overlapping content may also contribute to high alpha.

Sample quality and score interpretation

The package covers all deduplicated, gradable sittings logged from 30 July to 9 September 2026 across 16 locales. Of 369,429 result-page records, 1,611 same-visitor, identical-answer repeats were removed, leaving 367,818 sittings. No further records were excluded as incomplete or out of range. No time-based filter was applied; median completion time was 2 minutes 49 seconds in the 240,565 timed sittings.

These are sittings rather than verified unique people. The 354,438 sittings with an identifier, represent 337,172 distinct identifiers; 15,185 identifiers have more than one sitting. A further 13,380 sittings lack an identifier.

Straight-line responses were retained in 7,060 sittings (1.92%). They include 5,313 all-maximum and 1,188 all-minimum patterns, contributing to endpoint spikes. Such patterns may include genuine responses, so automatic exclusion is not warranted.

Score distribution

The overall mean is 57.9/100, the median is 59, and the standard deviation is 22.7. Skewness is -0.26 and excess kurtosis is -0.62 . The middle half of sittings falls between 42 and 75, while the 5th and 95th percentiles are 18 and 93. This broad spread is encouraging for differentiating self-reported experiences within this sample.

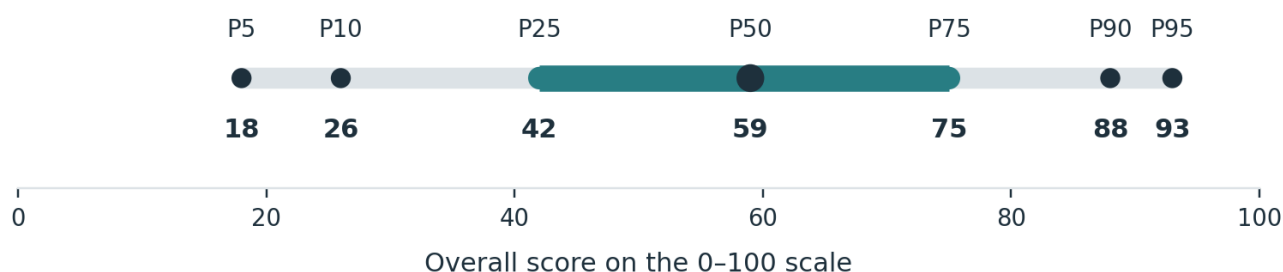


Figure 2. Reported overall-score percentiles. The teal segment marks the middle 50% of sittings; this is a quantile summary, not a reconstructed histogram. [1]

Scoring and reference values

All items use five response options stored as 0–4. Facets are item sums divided by 32 and multiplied by 100; the overall score uses the 24-item sum divided by 96 and multiplied by 100. Each score receives one final rounding to a whole number. A score of 75 means 75% of the possible item total, not the 75th percentile. The percentiles above describe self-selected website sittings and should be labelled sample reference values rather than population norms.

Facet patterns and structural interpretation

The three facets cover understandable aspects of romantic jealousy: suspicious or worried thoughts, emotional reactions to perceived threats, and checking or controlling behavior. Emotional items are more strongly endorsed on average than behavioral items. This is a descriptive pattern in the supplied sample, not evidence that the facets share an interchangeable severity scale.

Facet	Items	Mean	SD	Alpha
Cognitive jealousy	8	58.8	27.3	.90
Emotional jealousy	8	69.8	24.8	.89
Behavioral jealousy	8	45.2	24.7	.86

Table 1. Facet means and standard deviations use the 0–100 scoring scale.



Figure 3. Reported Pearson correlations between facet scores across all sittings.

Facet correlations range from .65 to .74. Their substantial positive overlap is compatible with a broader jealousy dimension, while the correlations remain below unity. Together with the content differences and facet reliabilities, this makes the three-part profile plausible and worth retaining for further evaluation.

The available summaries do not establish three distinct latent factors or show how much reliable facet-specific information remains after accounting for shared jealousy. First-principal-component loadings describe a broad component; they cannot substitute for a factor analysis of the proposed structure.

The next structural analysis should compare a one-factor model with three correlated factors using an estimator appropriate for ordinal responses. Omega estimates and, where justified, a general-factor model would help determine whether the overall score and each facet support the intended interpretations. These analyses should be checked in held-out data and within major language groups.

Item performance and response format

All 24 corrected item–total correlations are positive, ranging from .39 to .70; 23 items fall between .51 and .70. First-component loadings range from .42 to .75. These results are encouraging and reveal no item with a negative relationship to the scored total. The reported item–total statistics concern the overall scale, so within-facet item analyses are still needed.

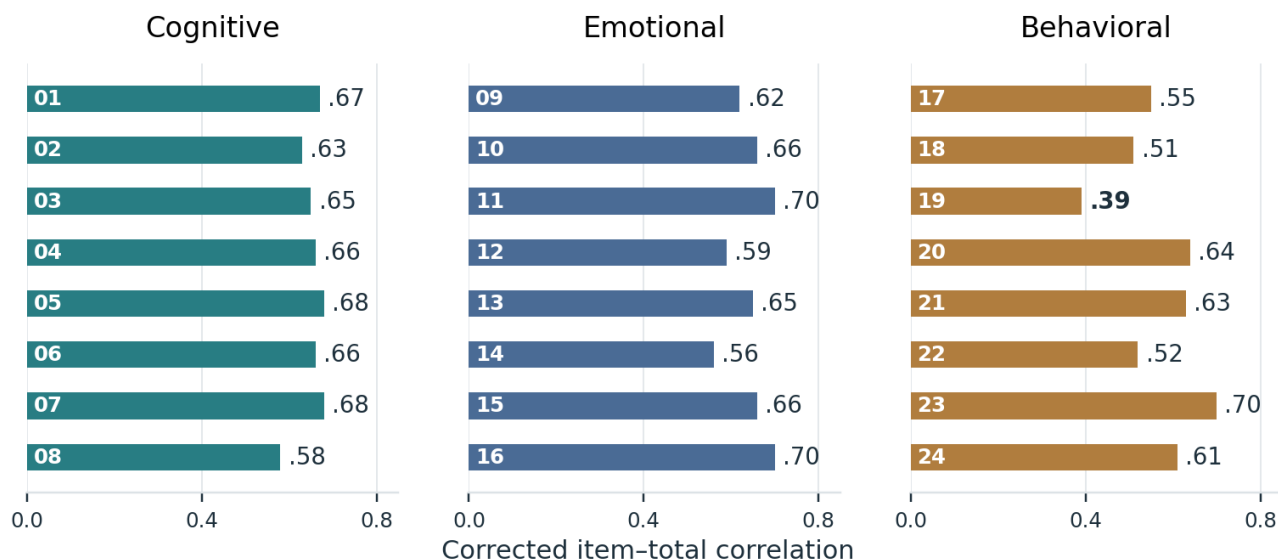


Figure 4. Corrected item–total correlations, grouped by assigned facet.

A targeted item review

Item 19, “Question my partner about their previous romantic relationships,” has the lowest item–total correlation (.39) and first-component loading (.42). Its wording may capture ordinary curiosity or relationship communication as well as jealous monitoring. It is a sensible candidate for cognitive interviewing and wording trials, rather than automatic deletion: its contribution to content coverage and within-facet performance should be examined first.

Several items refer to members of the “opposite sex.” Review should consider whether this phrasing works as intended across sexual orientations and relationship contexts. Any inclusive revision should preserve the intended construct and be tested for comprehension and comparability.

Validation of the adapted format

The test draws on Pfeiffer and Wong’s multidimensional framework. The original author-hosted scale uses seven response categories, whereas this IDRlabs version uses five-point agreement responses and adapted wording. Published evidence for the original instrument therefore provides a rationale for development, but does not automatically validate this implementation.

The adapted version merits direct comparison with the established scale, including its response instructions. Because all current items are keyed in the same direction, response-style sensitivity is also worth examining; adding reverse-worded items should not be treated as a simple remedy without testing.

Validation priorities and conclusion

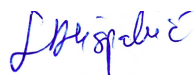
The size of this dataset makes a substantial next validation stage feasible. The following priorities would most directly strengthen confidence in the current score interpretations.

Next analyses

1. Establish robustness and structure. Recompute results using one sitting per identified respondent, examine unidentified sittings separately, and test reasonable completion-time and response-pattern exclusions. Compare ordinal factor models and report reliability beyond alpha.
2. Examine language and device differences. Russian accounts for 67% of sittings, Spanish 21%, and English 8%. Among the 241,550 device-identified sittings, 96% are mobile; mobile and desktop means are 60.9 and 52.2. The 8.7-point gap warrants investigation, but does not establish a device effect. Translation review, subgroup analyses, and measurement-invariance tests should precede strong cross-group comparisons.
3. Test external validity and stability. Compare scores with established jealousy measures, related relationship constructs, and relevant behavioral outcomes. Examine whether the facets add information beyond the overall score. A planned test–retest study should use a defined interval; spontaneous repeat sittings alone are not a test–retest design.

Conclusion

The Jealousy Test has a promising internal psychometric foundation. Strong consistency across all three facets, positive item discrimination, and broad score variation support its continued development and cautious educational use. The principal unresolved questions concern structural and external validity, the adapted response format, and generalization across languages and respondent groups. The supplied evidence does not establish diagnostic cut-offs or justify interpreting high scores as proof of pathological jealousy.



Sabina Alispahić, Ph.D.

Date: 30.9.2026.