

Psychometric Review of the Difficult Person Test

Jennifer Schulz, Ph.D., MSW, LSW

BSW Program Director | Associate Professor
Ashland University

21 August 2026

I conducted a preliminary psychometric review of the IDRIabs Difficult Person Test using datasets comprising 35 items and 187,801 valid test sessions. The purpose was to evaluate data quality, score distribution, internal consistency, item behaviour, and the functioning of the seven proposed facets: Callousness, Grandiosity, Aggression, Suspicion, Manipulativeness, Dominance, and Risk-taking.

1. Data quality

The original dataset contained 188,540 records. After duplicate sittings were identified and removed, 187,801 valid sessions remained. A total of 739 duplicate records were removed. No sessions were excluded for incomplete or malformed responses.

The available data include individual responses to all 35 items, seven facet scores, overall scores, test language, and limited geographic and technical information where available. Respondents are identified through a pseudonymous visitor identifier where available, and no names or contact information are contained in the dataset.

A total of 1,394 respondents, approximately 0.74% of the valid sample, selected the same response option for all 35 items. These records were retained in the supplied dataset. Such patterns produce isolated peaks at several exact overall scores, particularly 0, 50, and 100.

The sample is exceptionally large for an initial psychometric evaluation. However, respondents are self-selected internet users and should not be regarded as a representative normative sample of the general population. The linguistic composition is also uneven, with Russian-language respondents accounting for a substantial majority of the sample. This should be considered when developing norms or making cross-cultural interpretations.

2. Score distribution

The overall score distribution has the following characteristics:

- Mean = 49.23/100
- Median = 48.6
- SD = 18.45
- Skewness = 0.16
- Excess kurtosis = -0.22

The distribution is therefore broadly symmetrical, with only slight positive skew and somewhat less concentration around the centre than would be expected under a normal distribution. Scores extend across the full available range and show substantial variation among respondents.

The facet means vary considerably:

- Callousness: 41.9
- Grandiosity: 39.0
- Aggression: 61.7
- Suspicion: 57.2
- Manipulativeness: 53.2
- Dominance: 39.5
- Risk-taking: 52.3

Aggression has the highest average score, whereas Grandiosity and Dominance have the lowest averages. This variation indicates that respondents are not simply generating uniform response profiles across all seven proposed dimensions.

Several isolated peaks in the total-score distribution correspond to straight-line response patterns. Of the 1,394 straight-line responders, 943 obtained an overall score of 100, 226 obtained 0, and 179 obtained 50. Because these respondents constitute less than 1% of the valid sample, they do not materially determine the broader shape of the distribution.

3. Internal consistency

The complete 35-item instrument demonstrates very high internal consistency:

Cronbach's α = .931

After exclusion of straight-line responders:

Cronbach's α = .928

The small difference indicates that the very high reliability of the overall instrument is not dependent on straight-line responding. Such respondents appear to increase the coefficient only marginally.

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

- Callousness: α = .765
- Grandiosity: α = .891
- Aggression: α = .775
- Suspicion: α = .695
- Manipulativeness: α = .686
- Dominance: α = .787
- Risk-taking: α = .667

Grandiosity demonstrates particularly strong internal consistency for a five-item scale. Callousness, Aggression, and Dominance also show good reliability for brief subscales.

Suspicion and Manipulativeness show more moderate internal consistency, while Risk-taking has the lowest coefficient. An alpha of approximately .67 is not necessarily inadequate for a brief exploratory scale, but it suggests greater heterogeneity among its five items and warrants further examination.

Overall, however, none of the seven facets displays the pronounced reliability problem that would indicate an obviously dysfunctional subscale.

4. Item behaviour

All 35 items show positive corrected item-total correlations and therefore contribute in the expected direction to the overall Difficult Person score.

Across the complete instrument, corrected item-total correlations range from approximately $r = .29$ to $r = .66$.

Most items fall comfortably within a useful positive range. The strongest corrected item-total correlations occur for Item 28, at approximately $r = .66$, and Item 5, at approximately $r = .65$.

The weakest item is Item 31 in the Risk-taking facet, with a corrected item-total correlation of approximately $r = .29$. Items 35, 16, and 21 are also comparatively weaker, with correlations of approximately $r = .34$, $.35$, and $.35$, respectively.

These values merit attention but do not presently indicate serious scoring problems. In particular, removing Item 31 changes overall reliability only trivially, from approximately $\alpha = .9315$ to $\alpha = .9318$. Within the Risk-taking facet itself, removing Item 31 actually reduces reliability from approximately $\alpha = .667$ to $\alpha = .652$.

Item 4 behaves somewhat less consistently with the remaining Callousness items than most members of that facet. Removing it increases Callousness reliability from approximately $\alpha = .765$ to $\alpha = .770$. The improvement is very small and would not, by itself, constitute a strong psychometric reason for removing the item.

Importantly, there are no negative corrected item-total correlations. The data therefore provide no indication of incorrectly keyed items or of items operating in a direction opposite to the intended total score.

Items 31 and 35 may nevertheless benefit from review for wording and conceptual fit because Risk-taking is the least internally consistent of the seven facets.

5. Relationship between the facets

The relationships among the seven facets reveal a coherent but clearly multidimensional instrument.

All seven facets correlate positively with one another. Pairwise correlations range approximately from $r = .32$ to $r = .71$.

Some of the strongest relationships are:

- Callousness and Dominance: $r \approx .71$
- Callousness and Manipulativeness: $r \approx .66$
- Manipulativeness and Dominance: $r \approx .65$
- Aggression and Dominance: $r \approx .63$
- Grandiosity and Dominance: $r \approx .60$

The weakest relationships include:

- Grandiosity and Suspicion: $r \approx .32$
- Grandiosity and Risk-taking: $r \approx .35$
- Grandiosity and Aggression: $r \approx .39$

This pattern is consistent with the proposition that the seven dimensions share a broader underlying tendency while retaining meaningful differences from one another.

Dominance appears particularly strongly integrated with the wider instrument. Its relationship with the combined remainder of the test is stronger than that of any other facet. Grandiosity and Risk-taking appear somewhat more distinctive, although both remain positively associated with the other dimensions.

The reliability analysis provides additional support for this interpretation. Cronbach's alpha for the complete 35-item measure is approximately .931. Removing any one of the seven facets produces a lower coefficient rather than improving overall reliability. Alpha for the remaining 30 items ranges from approximately .915 to .928 depending on which facet is excluded.

There is therefore no evidence that any single facet is psychometrically working against the overall scale in the manner that would be expected from a substantially unrelated or incorrectly scored dimension.

At the same time, positive correlations and a high overall alpha do not demonstrate that the seven facets are interchangeable or that they necessarily constitute a strictly unidimensional construct. The pattern is compatible with a hierarchical structure in which several distinguishable interpersonal characteristics share a broader common component.

6. Construct interpretation

The current findings provide encouraging initial psychometric evidence for the Difficult Person Test.

The complete instrument demonstrates very high internal consistency, a broad overall score distribution, and generally positive item discrimination. All seven proposed facets are positively related to one another, and no facet appears to undermine the coherence of the overall measure.

The findings are therefore consistent with the existence of a broader dimension underlying Callousness, Grandiosity, Aggression, Suspicion, Manipulativeness, Dominance, and Risk-taking.

However, internal consistency does not establish construct validity.

The term "difficult person" encompasses a broad range of interpersonal characteristics, and the present dataset does not contain external psychological measures that would establish precisely what the overall score represents. The available data likewise do not establish the degree to which the measure converges with validated measures of antagonism, aggression, interpersonal dysfunction, narcissistic characteristics, suspiciousness, manipulativeness, impulsivity, or related personality constructs.

The high overall alpha should also be interpreted cautiously. A 35-item instrument composed of several correlated facets can produce a very high alpha even when those facets remain meaningfully distinguishable. Factor-analytic evidence is therefore necessary before concluding that the seven facets can be represented adequately by a single latent dimension.

Further validation should include:

- exploratory and confirmatory factor analysis of the seven proposed facets;
- comparison of a single-factor model with correlated-factor and hierarchical models;
- further examination of the Risk-taking facet, particularly Items 31 and 35;
- review of Item 4 within the Callousness facet;
- convergent validity against established measures of antagonism, aggression, interpersonal dysfunction, and related personality characteristics;
- discriminant validity against theoretically distinct psychological constructs;
- criterion validity using relevant interpersonal and behavioural outcomes;
- test-retest reliability;
- examination of measurement invariance across languages, countries, and relevant demographic groups;
- development of appropriate norms using more representative samples;
- continued investigation of straight-line and other potentially inattentive response patterns.

Cross-language validation is especially important because the current dataset contains responses from numerous language versions but is heavily weighted toward particular locales. Equivalent reliability in the pooled dataset does not by itself establish that individual items or facets function equivalently across translations.

7. Overall conclusion

The initial psychometric evidence for the Difficult Person Test is promising.

The complete 35-item instrument demonstrates very high internal consistency and substantial variation across respondents. All 35 items correlate positively with the remainder of the instrument, and the seven proposed facets form a coherent network of positively associated dimensions.

Grandiosity demonstrates particularly strong reliability as a brief five-item scale, while Callousness, Aggression, and Dominance also perform well. Suspicion and Manipulativeness show somewhat more moderate reliability, and Risk-taking is the weakest facet, although its reliability remains potentially usable for a brief exploratory measure.

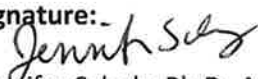
Unlike a scale containing a clearly anomalous or oppositely functioning dimension, no facet in the present instrument substantially detracts from the overall score. Removing any one facet reduces rather than increases the reliability of the remaining instrument.

The principal psychometric questions therefore concern dimensionality rather than an obvious scoring defect. The available evidence supports a broad common element across the seven facets, but it does not yet establish that this common element should be interpreted as a single unitary psychological construct.

The exceptionally large dataset provides a strong foundation for the next stage of validation. Factor analysis, external validity studies, test-retest assessment, and cross-language measurement-invariance testing would substantially strengthen the evidence base.

At present, the Difficult Person Test is best regarded as a multidimensional self-report measure of related interpersonal and personality characteristics, with a highly reliable overall composite score. Further construct-validation work is required before the total score should be interpreted as a definitive or clinically meaningful measure of a single underlying "difficult person" trait.

Signature:



Jennifer Schulz, Ph.D., MSW, LSW
Ashland University

21 August 2026