Psychometric characteristics of the MaxieDISC® questionnaire

Reliability, internal structure and measurement equivalence

Methodological report

1. Summary

This report presents an analysis of the psychometric properties of the MaxieDISC® questionnaire — a forced-choice instrument (24 blocks of 4 statements) measuring four dimensions of behavioural style: Dominance (D), Influence (I), Steadiness (S) and Conscientiousness (C). The analyses covered internal consistency, temporal stability (test–retest), internal structure (correlational, exploratory and confirmatory, including a Thurstonian IRT model) and measurement equivalence across sex. The data come from the years 2016–2026. The internal consistency of the scales, estimated with measures appropriate for ordinal data (ordinal alpha, McDonald's omega), ranges from 0.75 to 0.83; the split-half reliability of the estimates from the Thurstonian IRT model ranges from 0.81 to 0.87. The inter-correlations of the dimensions reproduce the theoretical structure of the model (strong negative associations of the opposing pairs D–S and I–C). The Thurstonian IRT model, appropriate for the forced-choice format, achieves good fit (RMSEA 0.041–0.042). Differential item functioning indices across sex are low.

2. The instrument

MaxieDISC® is a self-report questionnaire in a forced-choice format. In each of the 24 blocks the respondent ranks 4 statements from the most to the least self-descriptive. The scoring assigns statements to four dimensions (D, I, S, C) and aggregates the choices into a profile. The analyses were conducted on the natural profile.

3. Method

3.1. Samples

Table 1. Samples used in the individual analyses. Date range: 2016–2026.

Analysis	Sample (n)	Notes
Internal consistency	14,021	full item-level responses, main instance
Temporal stability (test–retest)	447 people / 299 pairs	returning respondents, 12 instances, intervals 1 mo–10 yr
Structure (EFA, correlations)	14,021	item-level responses
Thurstonian IRT model (fit + split-half reliability)	2,000	random sample
Equivalence (DIF, sex)	11,076	sex inferred from first name

3.2. Analyses and software

Internal consistency was estimated with classical Cronbach's alpha, ordinal alpha and McDonald's total omega (both on polychoric correlations), and additionally with the split-half reliability of the trait estimates from the Thurstonian IRT model (the correlation of estimates from two halves of the questionnaire, with the Spearman–Brown correction). For ordinal responses, ordinal alpha and omega are appropriate, and for the forced-choice format — the estimates from the Thurstonian model. Temporal stability was assessed with the intraclass

correlation coefficient ICC(2,1) and with the agreement of the dominant-style classification (percent agreement and Cohen's kappa). A profile was regarded as clear when the spread of scores between the extreme dimensions was at least 30 percentage points and the advantage of the dominant dimension over the next one was at least 10 percentage points. The internal structure was examined with analysis of inter-correlations, exploratory factor analysis, confirmatory factor analysis and the Thurstonian IRT model (WLSMV estimator), appropriate for forced-choice data. Measurement equivalence was assessed with differential item functioning (DIF) analysis across sex. Computations were carried out in R (packages lavaan, thurstonianIRT) and Python (factor_analyzer, statsmodels).

4. Results

4.1. Internal consistency

Classical Cronbach's alpha assumes interval data; for ordinal responses (ranks) ordinal alpha (based on polychoric correlations) and McDonald's omega are more appropriate. For the forced-choice format, the split-half reliability of the estimates from the Thurstonian IRT model is additionally appropriate. All four measures are reported below.

Table 2. Internal consistency of the scales (n=14,021; IRT split-half reliability: n=2,000).

Dimension	classical alpha	ordinal alpha	McDonald's omega	IRT split-half rel.
D	0.767	0.823	0.828	0.813
I	0.709	0.796	0.798	0.866
S	0.667	0.750	0.755	0.833
C	0.708	0.829	0.833	0.810

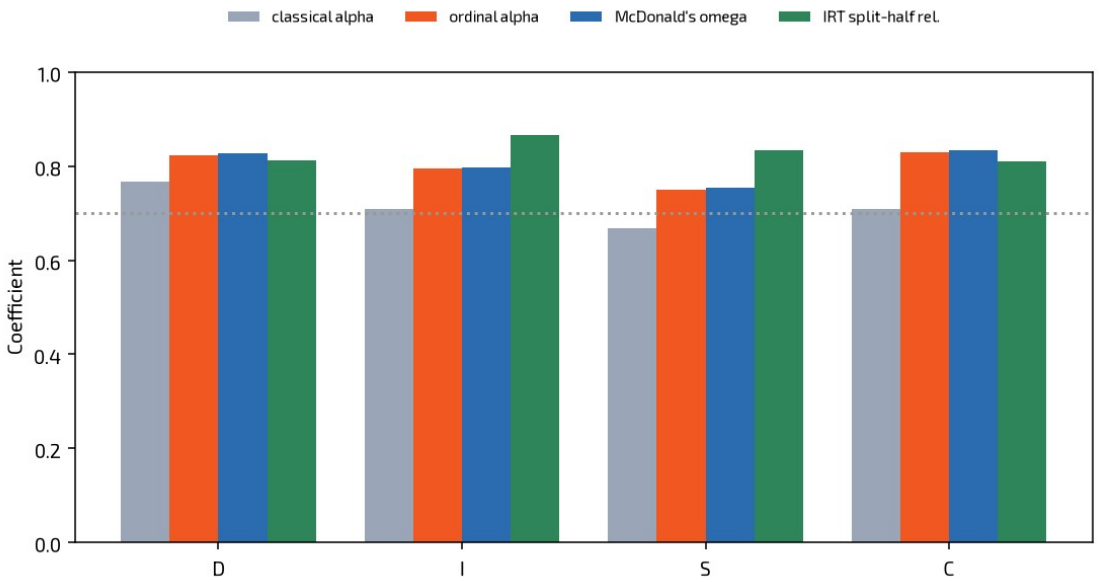


Figure 1. Internal consistency of the scales – comparison of four measures.

For ordinal data, ordinal alpha and omega range from 0.75 (S) to 0.83 (C), and the split-half reliability of the estimates from the Thurstonian IRT model ranges from 0.81 (C) to 0.87 (I). These values are higher than classical alpha (0.67–0.77), which underestimates reliability for ordinal data.

4.2. Temporal stability (test–retest)

Table 3. Temporal stability by inter-assessment interval (base sample).

Interval	n	ICC D	ICC I	ICC C	ICC S	Style agree.	κ style
1-3 mo	50	0.60	0.62	0.72	0.34	42%	0.11
3-6 mo	37	0.56	0.54	0.58	0.71	57%	0.32
6-12 mo	60	0.53	0.66	0.66	0.55	58%	0.42
1-2 yr	75	0.66	0.73	0.82	0.47	55%	0.33
2+ yr	104	0.61	0.77	0.82	0.62	64%	0.46

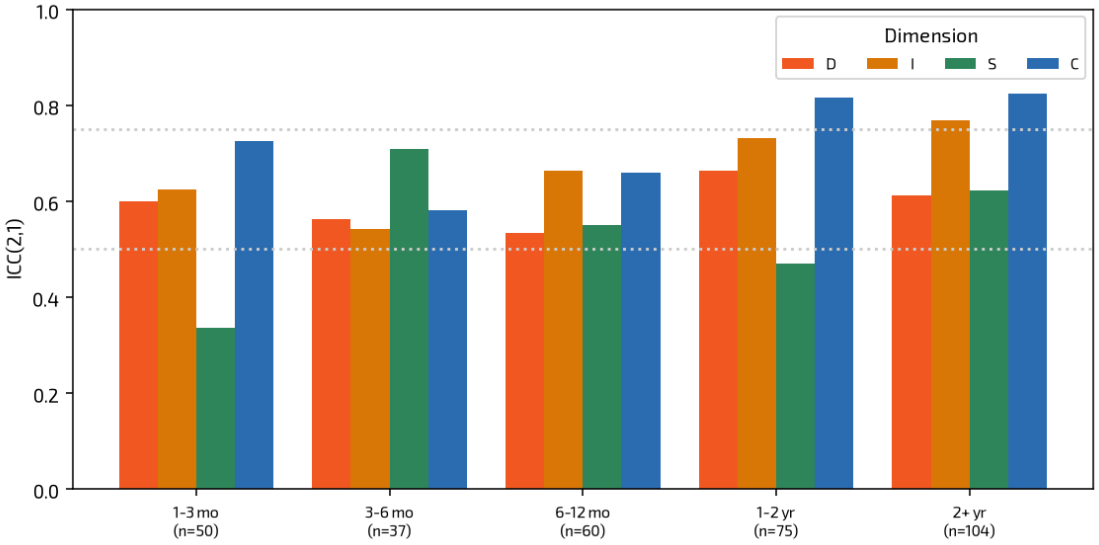


Figure 2. Test–retest reliability (ICC) by inter-assessment interval.

For longer intervals (1–2 and 2+ years) the ICC of dimensions C and I reaches 0.73–0.82 (Figure 2). The agreement of the dominant-style classification is 42–64% in the base sample and 50–75% after restricting to clear profiles (Figure 3).

Chance-corrected agreement (Cohen's κ) ranges from 0.11 to 0.46, indicating weak-to-moderate stability of the categorical classification. Continuous scores (ICC) are more stable than assignment to a single dominant style, which is typical of typological instruments when dimension values are close.

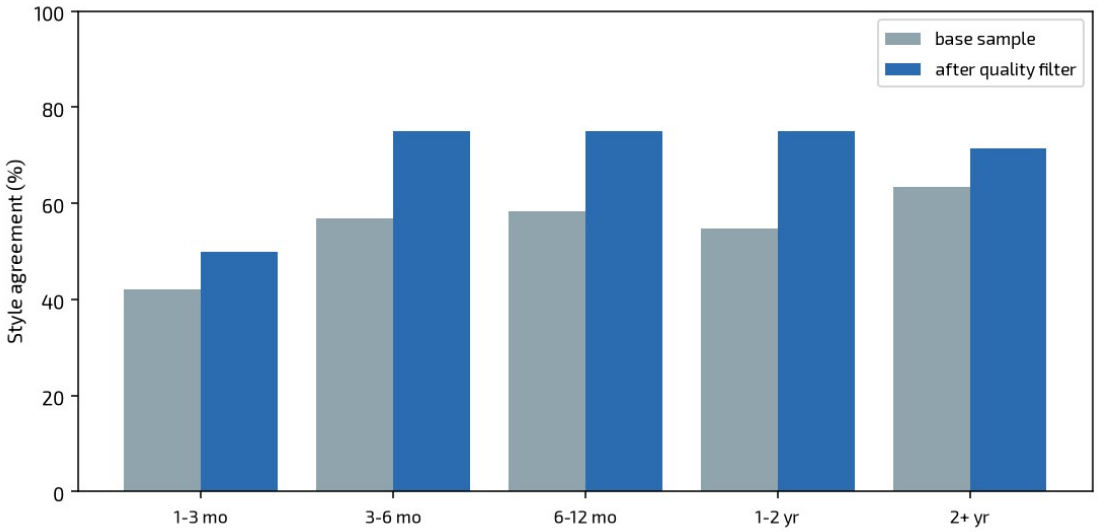


Figure 3. Agreement of the dominant-style classification by interval, in the base sample and after the profile-quality filter.

4.3. Internal structure

Table 4. Inter-correlations of the dimensions (n=14,021).

	D	I	S	C
D	+1.00	+0.14	−0.68	−0.49
I	+0.14	+1.00	−0.33	−0.70
S	−0.68	−0.33	+1.00	+0.20
C	−0.49	−0.70	+0.20	+1.00

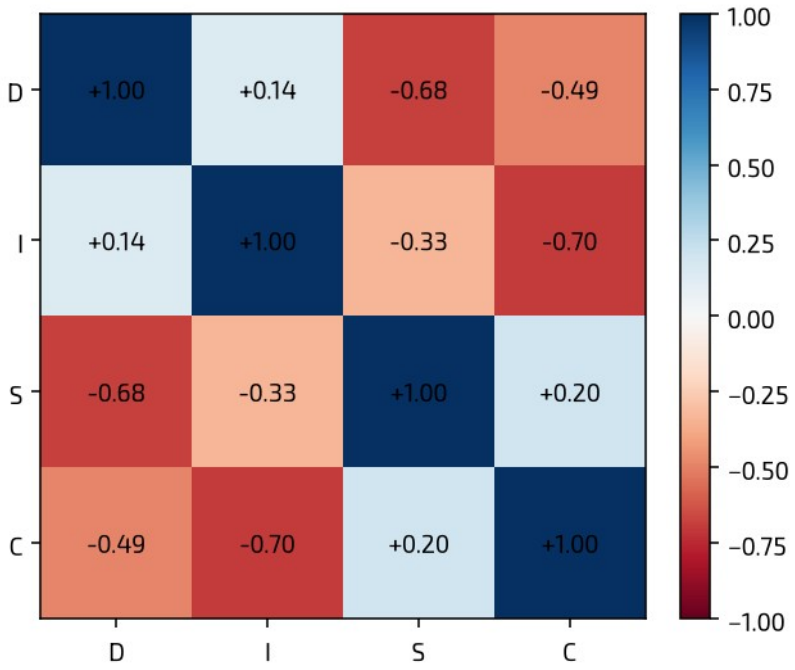


Figure 4. Inter-correlations of the dimensions (heatmap).

The strongest negative associations occur between the opposing pairs of the model: D–S (−0.68) and I–C (−0.70).

Exploratory factor analysis points to several dominant factors (eigenvalues: 9.1; 4.9; 4.3; 2.7; the remaining below 2 – Figure 5).

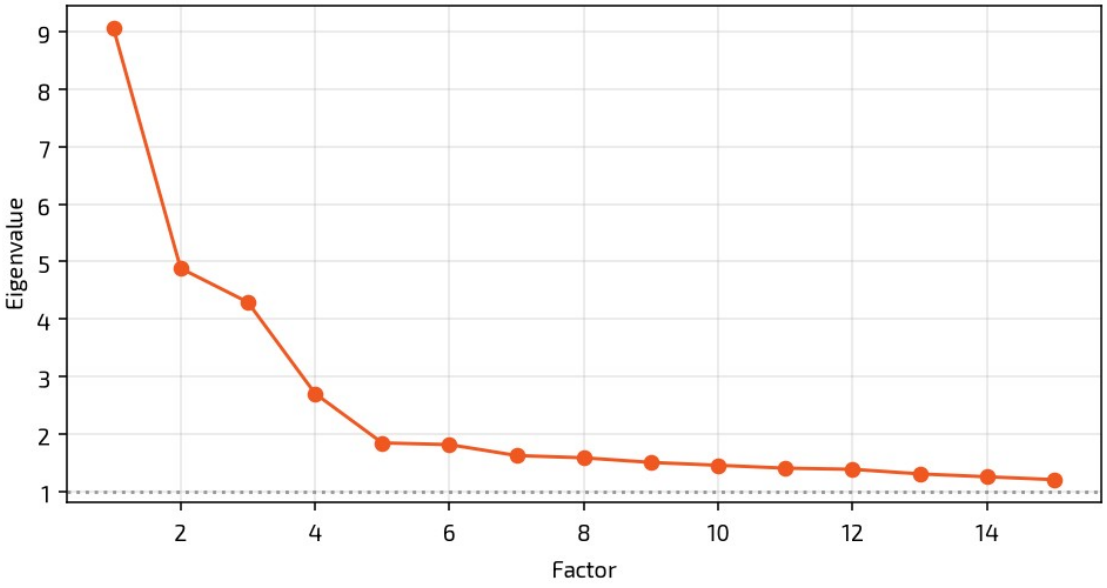


Figure 5. Scree plot (EFA, item-level responses).

Confirmatory factor analysis in the classical model achieves poor fit (CFI 0.38; RMSEA 0.087), consistent with the properties of forced-choice data. The Thurstonian IRT model, which accounts for the block structure of comparisons, achieves good fit for both halves of the questionnaire (RMSEA 0.041 and 0.042; Figure 6).

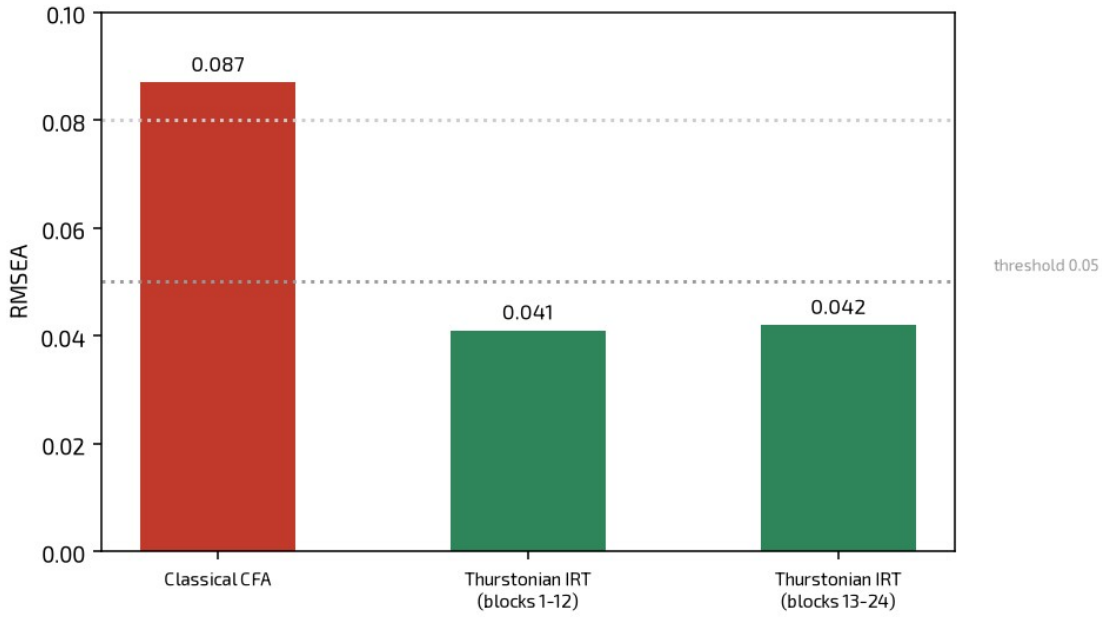


Figure 6. Fit (RMSEA) of the classical confirmatory factor analysis and the Thurstonian IRT model.

Table 5. Fit indices of the structural models (Thurstonian model: n=2,000).

Model	RMSEA	CFI	TLI
Classical CFA (4-factor)	0.087	0.38	0.36
Thurstonian IRT – blocks 1–12	0.041	0.875	0.869
Thurstonian IRT – blocks 13–24	0.042	0.853	0.846

4.4. Measurement equivalence (DIF, sex)

Table 6. Differential item functioning across sex (n=11,076).

Dimension	Items with DIF (/24)
D	4
I	0
S	0
C	1

In the I and S scales no items with significant differential functioning were found; in the D scale this concerns 4 of 24 items, and in the C scale – 1 of 24.

5. Summary of results

The MaxieDISC® scales show internal consistency (ordinal alpha, omega, split-half reliability of IRT estimates) in the range 0.75–0.87 and temporal stability that does not decline at longer intervals (ICC 0.53–0.82 for D, I and C). The inter-correlation structure corresponds to the theoretical four-dimension model with the opposing pairs D–S and I–C. The Thurstonian IRT model, appropriate for the forced-choice format, achieves good fit (RMSEA ≈ 0.04). Differential item functioning across sex is low.

6. Limitations

The data come from assessments conducted in natural settings (without experimental control of the interval and test–retest conditions). Internal consistency and structure were assessed on forced-choice data, for which classical factor methods yield biased results (hence the use of the Thurstonian IRT model). The Thurstonian model was estimated on a sample of n=2,000 and split into two halves of the questionnaire because of computational resources. Ordinal alpha and omega estimate the reliability of latent continuous variables rather than raw scores, and neither measure models the dependence among statements within a block (cf. Chalmers, 2018); this limitation is addressed by the Thurstonian IRT model. In each test–retest interval a different group of people was examined, so the ICC differences between intervals should not be interpreted as a trend over time. Sex in the DIF analysis was inferred from the first name. Population norms and DIF analysis by age were not within the scope of this study.

7. References

Brown, A., Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. Educational and Psychological Measurement, 71(3), 460–502.

Chalmers, R. P. (2018). On misconceptions and the limited usefulness of ordinal alpha. Educational and Psychological Measurement, 78(6), 1056–1071.

McDonald, R. P. (1999). Test theory: A unified treatment. Lawrence Erlbaum Associates.

Zumbo, B. D., Gadermann, A. M., Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. Journal of Modern Applied Statistical Methods, 6(1), 21–29.

Methodological document. Analyses on anonymised data.