

Structural Validity and Temporal Stability of DynaStrat, an Ipsative Energy-Allocation Instrument: Evidence from 2,193 Executive Profiles

Patrick Dura¹

Daniel Bos¹
Marion Noulhiane²

Clément Poirer²

¹Stratelio, Niort, France

²CEA-NeuroSpin, Université Paris Cité, France

patrick@stratelio.ai

April 2026

Abstract

We present psychometric evidence for DynaStrat, an ipsative behavioral energy-allocation instrument grounded in analytical psychology (Jung, 1921). This forced-distribution instrument constrains respondents to allocate a fixed budget across 12 behavioral modes organized in a hierarchical structure: 2 Hemispheres → 4 Quadrants → 6 Dialectical Pairs → 12 Modes → 48 Descriptors (the present study focuses on mode- and quadrant-level scores). Using archival data from $N = 2,193$ executives assessed between 2020 and 2026, we examine: (a) structural validity through multiple approaches adapted for ipsative data, including PCA structure recovery, drop-one-variable CFA, and Clemans correction; (b) internal consistency via intra-mode Cronbach’s alpha ($k = 4$ items per mode); (c) temporal stability via test-retest ICC; (d) measurement invariance across gender ($n_{\text{male}} = 1,285$; $n_{\text{female}} = 907$); and (e) the Gender $\times$ Age interaction estimated by moderated multiple regression on continuous age (20–82 years), complemented by Bayes Factors. Results demonstrate adequate structural coherence, good-to-excellent temporal stability, and uniformly small gender differences (mean $|d| = 0.22$) that are remarkably stable across the adult age range: gender $\times$ age interactions are negligible in magnitude (median $|\beta| = 0.08$ score points per year of age; all $\Delta R^2 \leq .002$; all Holm-adjusted $p > .98$), and Bayes Factors provide strong evidence for the no-interaction model on every mode (median $\text{BF}_{01} = 31.4$, all ≥ 10), indicating that gender-related patterns in energy allocation do not vary meaningfully with age in this executive sample. We discuss implications for ipsative measurement methodology and the use of forced-distribution instruments in executive coaching and team development.

Keywords: ipsative assessment, forced-choice, psychometric validation, executive coaching, behavioral energy allocation, analytical psychology, test-retest reliability, gender differences

1 Introduction

Behavioral assessment in organizational settings relies heavily on normative instruments such as the Big Five Inventory (Goldberg, 1990), the Myers-Briggs Type Indicator (Myers & McCaulley, 1985), or DISC-based tools. These instruments measure traits in absolute terms—each dimension is scored independently, and an individual can theoretically score high on all dimensions simultaneously. While this approach offers interpretive simplicity, it presents several limitations in the context of executive coaching and leadership development. First, normative instruments are susceptible to acquiescence bias and socially desirable responding, particularly salient in high-stakes organizational contexts (Meade, 2004). Second, they treat behavioral dimensions as independent, which contradicts the practical reality that leaders operate under finite attentional and energetic resources: investing heavily in strategic vision necessarily means investing less in operational execution.

Ipsative instruments address these concerns by de-

sign. In a forced-distribution format, respondents must allocate a fixed budget across behavioral dimensions, making it impossible to endorse all dimensions equally. This eliminates acquiescence bias, substantially reduces socially desirable responding (Hicks, 1970), and produces profiles that reflect the *relative* allocation of behavioral energy—arguably a more ecologically valid representation of how individuals actually operate in professional contexts.

Despite these advantages, ipsative instruments have been criticized on psychometric grounds. Because scores are interdependent (they must sum to a constant), classical test theory assumptions are violated: observed correlations include artificial negative bias of $-1/(n - 1)$ where n is the number of ipsative components (Clemans, 1966), covariance matrices are singular, and standard reliability and factor analytic methods produce distorted results (Dunlap & Cornwell, 1994). These concerns led some authors to caution against the use of ipsative measures in applied settings (Baron, 1996).

Recent methodological advances have substantially

Table 1: Positioning of DynaStrat Relative to Established Personality / Style Instruments

Design dimension	DynaStrat	MBTI	Big Five / NEO-PI-R	DISC
Scoring metric	Ipsative (forced distribution)	Quasi-ipsative (dichotomous)	Normative (Likert)	Mostly normative (Likert)
Construct framing	Behavioral <i>energy allocation</i>	Cognitive type preferences	Latent traits (independent)	Behavioral style (interpersonal)
Theoretical grounding	Analytical psychology (Jung, 1921)	Jungian type theory (Myers & McCaulley, 1985)	Lexical / factor-analytic	Marston (1928); (Inscape Publishing, 2013)
Hierarchical resolution	5 levels: 2 / 4 / 6 / 12 / 48; 240 indicators total	4 dichotomies, 16 types	5 traits, 30 facets	4 styles

mitigated these concerns. The Thurstonian IRT framework (Brown & Maydeu-Olivares, 2011, 2012) provides a rigorous approach to modeling forced-choice data by estimating latent utilities that underlie the observed ranking process. Several correction methods for ipsative correlations have been proposed (Clemans, 1966; Chan & Bentler, 1993), and PCA-based approaches have been shown to validly recover latent structures from ipsative data (ten Berge, 1999). Recent large-scale validation studies have confirmed that forced-choice instruments can achieve adequate psychometric properties when analyzed with appropriate methods (Li et al., 2025).

1.1 Positioning Relative to Established Instruments

For readers familiar with the dominant instruments used in organizational assessment, Table 1 situates DynaStrat against MBTI, the Big Five / NEO-PI-R, and DISC along four design dimensions: scoring metric, construct framing, theoretical grounding, and hierarchical resolution.

Two contrasts are worth flagging because they are routinely conflated in the literature. First, the shared Jungian heritage of MBTI and DynaStrat does not imply design equivalence: MBTI uses a quasi-ipsative dichotomous format that yields independent type-preference scores, whereas DynaStrat enforces a strict fixed-sum constraint ($\sum_{m=1}^{12} x_m = T$) at the score level, producing a forced-distribution profile of relative allocations. Second, DynaStrat’s construct framing—energy allocation under a finite budget—differs in kind from trait or type framings, in that profiles are interpreted as *trade-offs* rather than independent levels (an executive cannot become “more strategic” on DynaStrat without becoming less something else). The implications of these design choices for bias resistance and ecological validity are discussed further in Hicks (1970); Meade (2004); Li et al. (2025).

1.2 The Present Study

In this paper, we present the psychometric properties of DynaStrat, an ipsative energy-allocation instrument used in executive coaching and leadership development since 2011. Using archival data from executive profiles collected over six years (2020–2026), we provide evidence for the instrument’s structural validity, internal consistency, temporal stability, and measurement

invariance across gender, and we test whether observed gender differences are stable across the adult age range via continuous-age moderated multiple regression with Bayesian inference for null findings.

2 The Instrument

2.1 Theoretical Foundations

DynaStrat is rooted in Jung’s theory of psychological types (Jung, 1921), which posits that individual differences in cognition and behavior reflect preferred modes of psychological functioning organized along complementary polarities. This theoretical framework is operationalized through a resource-allocation model in which behavioral energy is conceptualized as a finite budget distributed across multiple modes of functioning.

The central construct—behavioral energy allocation—reflects the observation that individuals systematically prioritize certain behavioral modes over others, and that this prioritization pattern constitutes a stable individual difference relevant to professional performance, leadership style, and team dynamics.

2.2 Hierarchical Architecture

The instrument measures behavioral energy across a five-level hierarchical structure:

- Level 5: 2 Hemispheres** — Exploration (openness, relationships, abstraction) versus Structure (analysis, organization, implementation).
- Level 4: 4 Quadrants** — Vision (strategic abstraction), Relation (interpersonal connection), Realization (operational execution), and Structuring (analytical organization), each formed by the intersection of two territorial axes.
- Level 3: 6 Dialectical Pairs** — Complementary polar tensions between modes, reflecting the fundamental trade-offs in energy allocation.
- Level 2: 12 Behavioral Modes** — The core measurement level, each representing a distinct pattern of behavioral energy investment.
- Level 1: 48 Descriptors** — Four behavioral descriptors per mode, each assessed on five facets,

DynaStrat Hierarchical Architecture

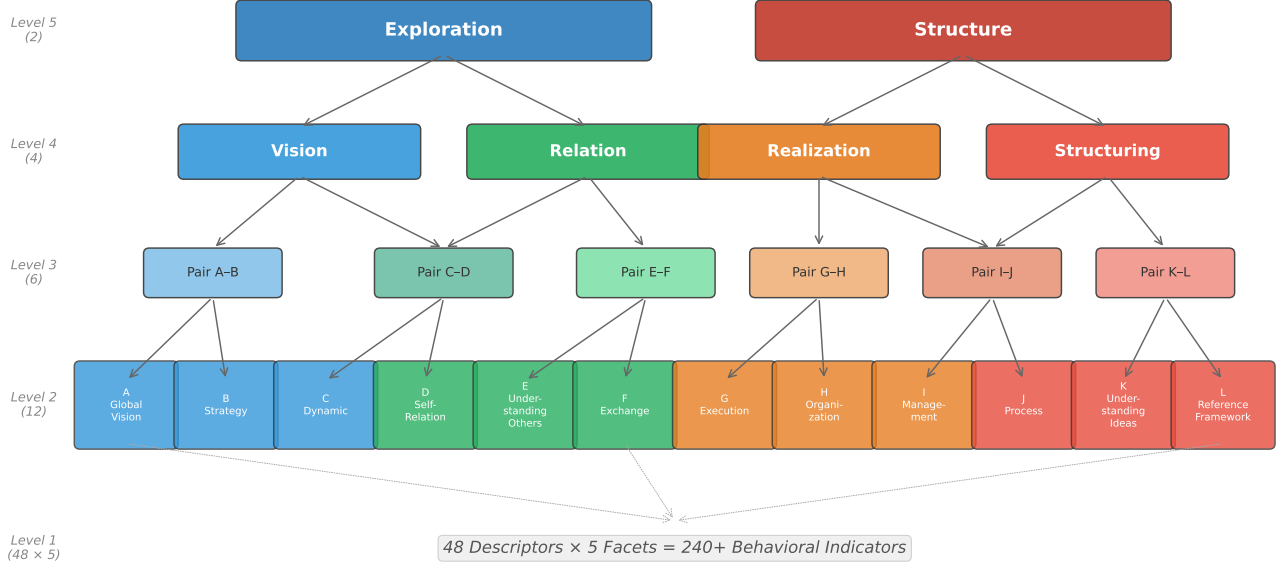


Figure 1: DynaStrat hierarchical architecture. The instrument measures behavioral energy allocation across five nested levels: 2 Hemispheres (Exploration vs. Structure), 4 Quadrants, 6 Dialectical Pairs (complementary tensions between adjacent modes: A–B, C–D, E–F, G–H, I–J, K–L), 12 Behavioral Modes, and 48 Descriptors (each assessed on 5 facets). Pairs C–D and I–J span the boundary between two quadrants. Arrows indicate compositional relationships.

producing over 240 behavioral indicators per profile.

The quadrant scores are composites of three mode scores each: Vision = Modes A + B + C; Relation = Modes D + E + F; Realization = Modes G + H + I; Structuring = Modes J + K + L. Figure 1 illustrates this hierarchical architecture and Table 2 describes the 12 modes and their quadrant membership.

2.3 Ipsative Mechanism

The instrument employs a forced-distribution scoring system in which the assessment procedure constrains each respondent’s scores to sum to a fixed total T across the 12 modes. Individual mode scores are derived from the aggregation of responses across the five assessment parts; the scoring algorithm maps item-level responses (rankings, binary choices, and selections) onto mode-level point allocations such that $\sum_{m=1}^{12} x_m = T$ for every respondent. This design ensures that each score reflects the *relative* investment in a behavioral mode rather than its absolute level. The forced-distribution constraint means that elevating one score necessarily reduces others, producing inherently interdependent scores. This ipsativity is a deliberate design feature, not a measurement artifact, as it captures the zero-sum nature of attentional and energetic resources in real-world

executive functioning.

2.4 Administration

The assessment comprises five parts combining item ranking, binary choices, and multi-selection tasks. In a typical task, respondents are presented with a set of behavioral descriptors (e.g., “anticipates long-term consequences,” “builds trust through active listening”) and must rank-order them according to how well each describes their habitual professional behavior. By requiring respondents to distribute a fixed number of points across descriptors, each task contributes to the forced-distribution constraint that produces the ipsative score profile. The assessment is administered individually online with a typical completion time of 25–40 minutes ($M = 36.7$ min, $SD = 20.3$). The instrument is available in seven languages (French, English, Spanish, Italian, German, Swedish, Norwegian).

3 Method

3.1 Participants

The sample comprised $N = 2,193$ executives and senior managers who completed DynaStrat V1 between 2020 and 2026 as part of executive coaching, leadership de-

Table 2: The 12 DynaStrat Behavioral Modes

Mode	Label	Quadrant	Behavioral Focus
A	Global Vision	Vision	Long-range, systemic perspective-taking
B	Strategy	Vision	Strategic planning and anticipation
C	Dynamic	Vision	Initiative, drive, and forward momentum
D	Self-Relationship	Relation	Introspection and self-awareness
E	Understanding Others	Relation	Empathy and interpersonal sensitivity
F	Exchange	Relation	Communication and relational reciprocity
G	Execution	Realization	Action-oriented task completion
H	Organization	Realization	Structuring work and coordinating resources
I	Management	Realization	Directing, delegating, and supervising
J	Process	Structuring	Analytical rigor and procedural discipline
K	Understanding Ideas	Structuring	Conceptual analysis and knowledge integration
L	Reference Framework	Structuring	Norms, standards, and theoretical benchmarking

velopment, or organizational assessment programs. After excluding internal test accounts, duplicate assessments (retaining the most recent per individual), extreme completion times (< 5 or > 180 minutes), implausible ages (< 18 or > 85 years), and V2 profiles ($n = 89$; see below), the final analytic sample included 2,193 unique profiles. Table 3 presents the demographic characteristics of the sample.

Table 3: Demographic Characteristics of the Sample ($N = 2,193$)

Variable / Category	n	%
<i>Gender</i>		
Male	1,285	58.6
Female	907	41.4
Not reported	1	< 0.1
<i>Age (years)</i>	$M = 46.5$ ($SD = 9.5$)	
Median		47
Range		20–82
<i>Age band</i>		
< 30	85	3.9
30–39	452	20.6
40–49	753	34.3
50–59	731	33.3
≥ 60	172	7.8

The sample is predominantly composed of mid-to-senior career executives (67.6% aged 40–59), reflecting the instrument’s primary use in executive coaching contexts. The gender distribution (58.6% male) is consistent with the representation of women in senior management roles in France during the assessment period. A revised version of the instrument (V2, $n = 89$) was deployed during the assessment window; V2 introduced wording changes to the descriptors in Modes D and L. Because formal psychometric equivalence between V1 and V2 has not yet been established, the present analyses are restricted to V1 profiles only. A dedicated V2 validation study is planned once sufficient sample size is reached.

3.2 Procedure

All data were archival, collected in the normal course of professional practice. Participants completed the instrument as part of coaching engagements, leadership development programs, or organizational assessments. No experimental manipulation was involved. All participants provided informed consent for the use of their anonymized data in research. Data were extracted from the operational database, anonymized, and analyzed using Python (pandas, scipy, semopy, pingouin).

3.3 Analytic Strategy

Given the ipsative nature of the data, we employed a multi-method approach to structural validation, as classical CFA and EFA methods produce distorted results when applied to ipsative scores (Clemans, 1966; Chan & Bentler, 1993). The following analyses were conducted:

Structural Validity. We tested six competing CFA models (unidimensional, 2-hemisphere, 4-quadrant, 6-pair, hierarchical, and a 4-quadrant-to-2-hemisphere variant) on the raw ipsative scores to establish a baseline comparison, then applied three ipsative-appropriate approaches: (a) Clemans correction (Clemans, 1966) to remove the artificial negative correlation bias before CFA; (b) drop-one-variable CFA (Chan & Bentler, 1993) to break the covariance matrix singularity; and (c) PCA structure recovery with varimax rotation (ten Berge, 1999) followed by Procrustes rotation toward the theoretical target matrix.

Internal Consistency. Cronbach’s alpha was computed *within* each mode on its four constituent descriptors. Because descriptors within a mode are not subject to the ipsative constraint (only mode-level scores are interdependent), alpha is a valid reliability estimate at this level. With only $k = 4$ items per mode, alpha is structurally constrained: the Spearman-Brown formula predicts that scales with $k = 4$ items and moderate inter-item correlations ($\bar{r} \approx .30$) yield $\alpha \approx .63$. Following recommendations for short scales (Loewenthal, 2001; Briggs & Cheek, 1986), $\alpha \geq .50$ was adopted as the acceptability threshold.

Temporal Stability. Test-retest reliability was assessed using intraclass correlation coefficients—ICC(3,1), two-way mixed model, consistency—on $n = 96$ individuals who completed the instrument twice with intervals ranging from 41 days to approximately 3 years. Cicchetti’s (1994) benchmarks were used: $\geq .75 = \text{excellent}$, $.60\text{--}.74 = \text{good}$, $.40\text{--}.59 = \text{moderate}$.

Measurement Invariance. Approximate multi-group CFA was conducted across gender, comparing fit indices between male and female subsamples, along with loading comparisons and Cohen’s d effect sizes. Criteria from Cheung & Rensvold (2002) and Chen (2007) were applied: $|\Delta\text{CFI}| \leq .01$ and $|\Delta\text{RMSEA}| \leq .015$.

Gender $\times$ Age Interaction. To examine whether gender differences in energy allocation are stable across the lifespan or vary with age and career stage, we used two complementary analyses. For visualization and comparison with the binned literature, Cohen’s d was computed within five age bands (< 30 , $30\text{--}39$, $40\text{--}49$, $50\text{--}59$, ≥ 60) for each mode and quadrant. For inference, however, we kept age continuous and fit a moderated multiple regression (Aiken & West, 1991) for each mode of the form $\text{score} = \beta_0 + \beta_1 \text{gender}_{\text{eff}} + \beta_2 \text{age}_c + \beta_3 (\text{gender}_{\text{eff}} \times \text{age}_c) + \varepsilon$, where gender was effects-coded ($M = +0.5$, $F = -0.5$) and age was mean-centred so that β_3 is identifiable and quantifies the change in the M–F gap per year of age. We report β , SE, 95% CI, t , p , and ΔR^2 for the interaction term, with Holm correction applied across the 12 interaction tests. Because the absence of a significant interaction in a frequentist framework does not by itself constitute evidence for invariance (Wagenmakers, 2007), we additionally computed two Bayes Factors per mode: BF_{10} for the gender main effect (Bayesian independent-samples t -test, default Cauchy prior with scale $r = \sqrt{2}/2$) and BF_{01} for the gender $\times$ age interaction via the BIC approximation $\text{BF}_{01} \approx \exp((\text{BIC}_{\text{full}} - \text{BIC}_{\text{main}})/2)$. Following Jeffreys’ conventions (Lee & Wagenmakers, 2014), $\text{BF}_{01} \geq 3$ was interpreted as moderate evidence for invariance and $\text{BF}_{01} \geq 10$ as strong evidence.

4 Results

4.1 Descriptive Statistics

Table 4 and Figure 2 present the descriptive statistics for the 12 mode scores. The fixed-sum constraint produces a theoretical grand mean per mode determined by the total budget divided across the 12 modes. Observed means ranged from 49.1 (Mode L) to 85.6 (Mode C), indicating systematic patterns of energy allocation across the executive population. Score distributions were approximately symmetric for most modes (skewness range: -0.39 to 0.72), with light tails (kurtosis range: -0.65 to 0.64). Shapiro-Wilk tests rejected normality for all modes ($p < .001$), as expected with $N > 2,000$.

At the quadrant level, Vision received the highest mean allocation ($M = 238.0$, $SD = 52.3$), followed by Relation ($M = 230.7$, $SD = 45.3$), Realization ($M = 204.1$, $SD = 47.7$), and Structuring ($M = 167.3$, $SD = 40.4$). This pattern—with exploration-oriented quadrants (Vision, Relation) receiving more energy than structure-oriented quadrants (Realization, Structuring)—is consistent with the executive population’s typical focus on strategic and relational leadership functions.

4.2 Structural Validity

4.2.1 Standard CFA on Raw Ipsative Scores

Standard CFA on the raw ipsative scores is presented in Table 6. The mean raw inter-mode correlation was $r = -.089$, matching the theoretical ipsative artifact of $-1/(n-1) = -1/11 = -.091$ almost exactly, confirming that the observed negative correlations are driven by the fixed-sum constraint rather than by substantive relationships. The covariance matrix was rank-deficient (smallest eigenvalue = 0.000), precluding maximum-likelihood EFA.

Despite the ipsative constraint, multi-factor models showed a clear hierarchical pattern of improving fit, with the 6-pair model (M4, CFI = .863) and 4-quadrant model (M3, CFI = .829) approaching acceptable thresholds. This level of fit—unusual for raw ipsative data—supports the structural coherence of the theoretical model. The strong differentiation between the unidimensional model (CFI = .476) and multi-factor models confirms that the data carry substantial structural information beyond the ipsative artifact.

4.2.2 Ipsative-Corrected Structural Analyses

Three approaches adapted for ipsative data converged on supporting the four-quadrant structure.

PCA Structure Recovery. Following ten Berge (1999), we applied PCA to the ipsative scores, which does not require a positive-definite covariance matrix. Four varimax-rotated components explained 79.9% of the total variance (eigenvalues: 5.69, 1.97, 1.23, 0.70). Parallel analysis (Figure 6) confirmed that three components exceeded the 95th percentile of random eigenvalues, while the fourth (0.70) fell just below the random threshold but loaded coherently on the theoretical Structuring quadrant. The varimax-rotated loadings (Figure 7) showed a block-diagonal pattern consistent with the theoretical quadrants, with the strongest modes loading at $|.6|$ – $|.9|$ on their hypothesized component. Procrustes rotation toward the theoretical four-quadrant target matrix (Figure 8) yielded a mean Tucker congruence coefficient of $\phi = .71$, with per-quadrant values of $\phi_{\text{Vision}} = .71$, $\phi_{\text{Relation}} = .76$, $\phi_{\text{Realization}} = .64$, and $\phi_{\text{Structuring}} = .73$. These values fall below the conventional “fair similarity” threshold of $\phi \geq .85$ (Lorenzo-Seva & ten Berge, 2006), which is unsurprising given the structural distortions that the

Table 4: Descriptive Statistics for 12 Mode Scores ($N = 2,193$)

Mode	M	SD	Mdn	$Skew$	$Kurt$	P_{10}	P_{50}	P_{90}
A (Global Vision)	72.7	21.2	72	0.02	-0.65	44	72	100
B (Strategy)	79.7	21.2	80	-0.07	-0.50	52	80	108
C (Dynamic)	85.6	20.1	86	-0.15	-0.49	58	86	112
D (Self-Relationship)	69.9	19.8	70	0.03	-0.63	44	70	96
E (Understanding Others)	75.6	19.3	76	-0.10	-0.35	50	76	100
F (Exchange)	85.1	17.3	86	-0.39	0.09	62	86	106
G (Execution)	73.3	19.6	72	0.03	-0.47	48	72	98
H (Organization)	68.2	19.2	68	0.12	-0.60	44	68	94
I (Management)	62.6	17.4	62	0.18	-0.41	40	62	86
J (Process)	56.7	20.5	56	0.34	-0.54	30	56	84
K (Understanding Ideas)	61.4	15.8	60	0.26	-0.20	42	60	82
L (Reference Framework)	49.1	14.5	48	0.72	0.64	32	48	68

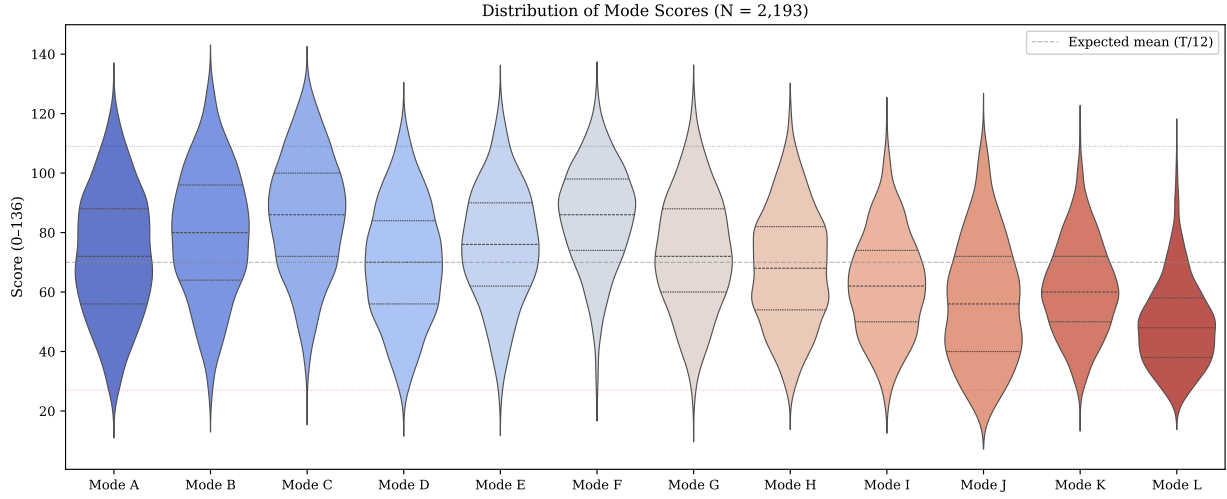


Figure 2: Distribution of scores across the 12 DynaStrat modes ($N = 2,193$). Violin widths represent kernel density estimates; horizontal lines within each violin indicate quartiles. Modes are colored by quadrant membership: Vision (blue, A–C), Relation (light blue, D–F), Realization (beige, G–I), Structuring (red, J–L). The dashed line indicates the expected mean under uniform allocation ($T/12$).

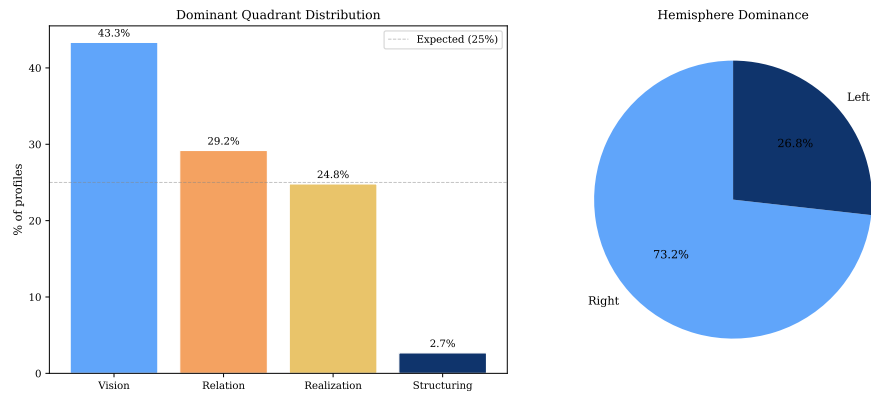


Figure 3: Distribution of quadrant scores ($N = 2,193$). Vision and Relation receive substantially more energy than Realization and Structuring, consistent with the executive sample’s strategic/relational focus.

Confirmatory Factor Analysis: Model Comparison

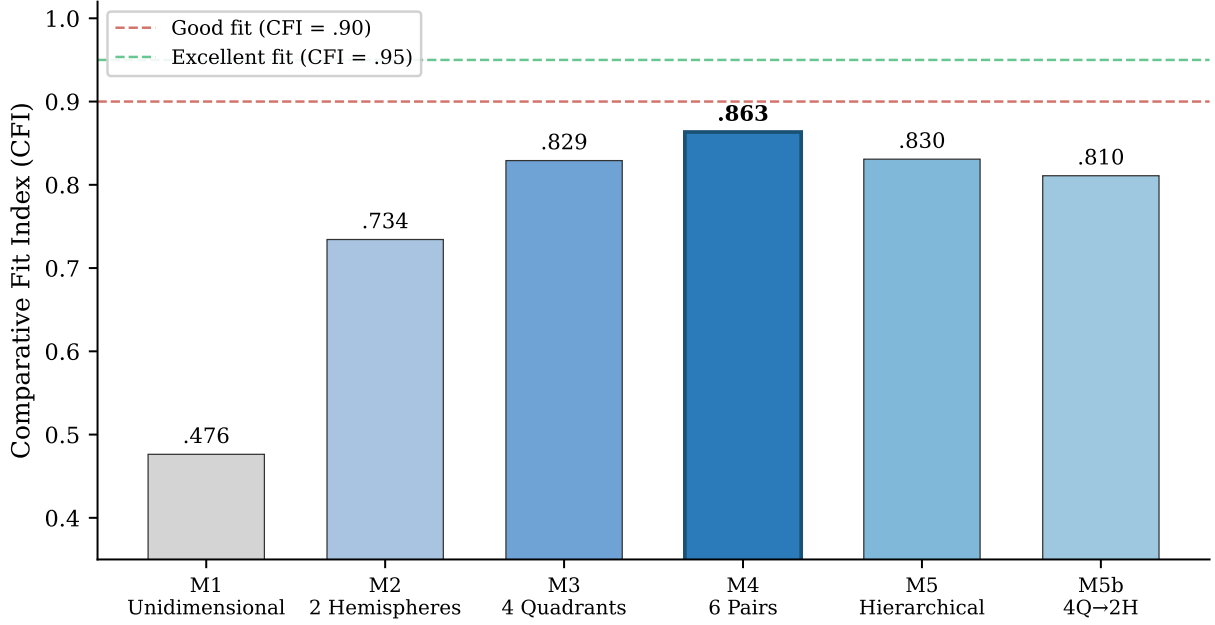


Figure 4: Comparative Fit Index (CFI) across six CFA models of increasing structural complexity. The 6-pair model (M4) achieved the highest CFI (.863). Dashed lines indicate conventional thresholds for good (CFI = .90) and excellent (CFI = .95) fit, which are structurally unattainable for raw ipsative data due to the singular covariance matrix.

Table 5: Descriptive Statistics for Quadrant Scores ($N = 2,193$)

Quadrant	M	SD	Mdn	P_{10}	P_{90}
Vision	237.9	52.4	239	168	308
Relation	230.5	45.4	230	172	288
Realization	203.9	47.7	202	140	268
Structuring	167.7	40.5	164	118	222

Table 6: CFA Model Comparison on Raw Ipsative Scores ($N = 2,193$)

Model	χ^2	CFI	RMSEA
M1: Unidimensional	12,283	.476	.315
M2: 2 Hemispheres	6,256	.734	.227
M3: 4 Quadrants	4,038	.829	.191
M4: 6 Pairs	3,229	.863	.189
M5: Hierarchical	3,990	.831	.208
M5b: 4Q→2H	4,464	.811	.199

Note. Degrees of freedom are omitted because the singular covariance matrix (rank = $n - 1 = 11$) renders standard df computation unreliable. Conventional thresholds (CFI $\geq .90$, RMSEA $\leq .06$) are not applicable to ipsative data: the fixed-sum constraint artifactually degrades all fit indices.

fixed-sum constraint imposes on the inter-mode covariance geometry. We therefore interpret the Procrustes results as evidence of *partial*—not full—recovery of the theoretical four-quadrant structure: the PCA-extracted components are aligned with the theoretical quadrants in pattern but not at the magnitudes that would license a strict “fair similarity” claim under L-S & ten Berge’s criteria, with the strongest alignment on Relation ($\phi = .76$) and the weakest on Realization ($\phi = .64$). The combined evidence from PCA loadings, parallel analysis, Clemans-corrected correlations, and drop-one CFA (below) converges on a four-quadrant interpretation, while acknowledging that no single ipsative-corrected method achieves the fit standards available to normative instruments.

Clemans-Corrected Correlations. Applying the Clemans (1966) correction to remove the artificial negative correlation bias, the corrected inter-mode correlation matrix (Figure 9) showed substantially clearer block-diagonal structure, with within-quadrant correlations averaging $r = +.40$ and between-quadrant correlations averaging $r = -.04$.

Drop-One-Variable CFA. Following Chan & Bentler (1993), we dropped one mode (L) and re-estimated the four-quadrant CFA on the remaining 11 modes. The resulting model achieved CFI = .883 and RMSEA = .102, both substantially improved over the full-set CFA, demonstrating that the singular covariance matrix accounts for a significant portion of the apparent misfit.

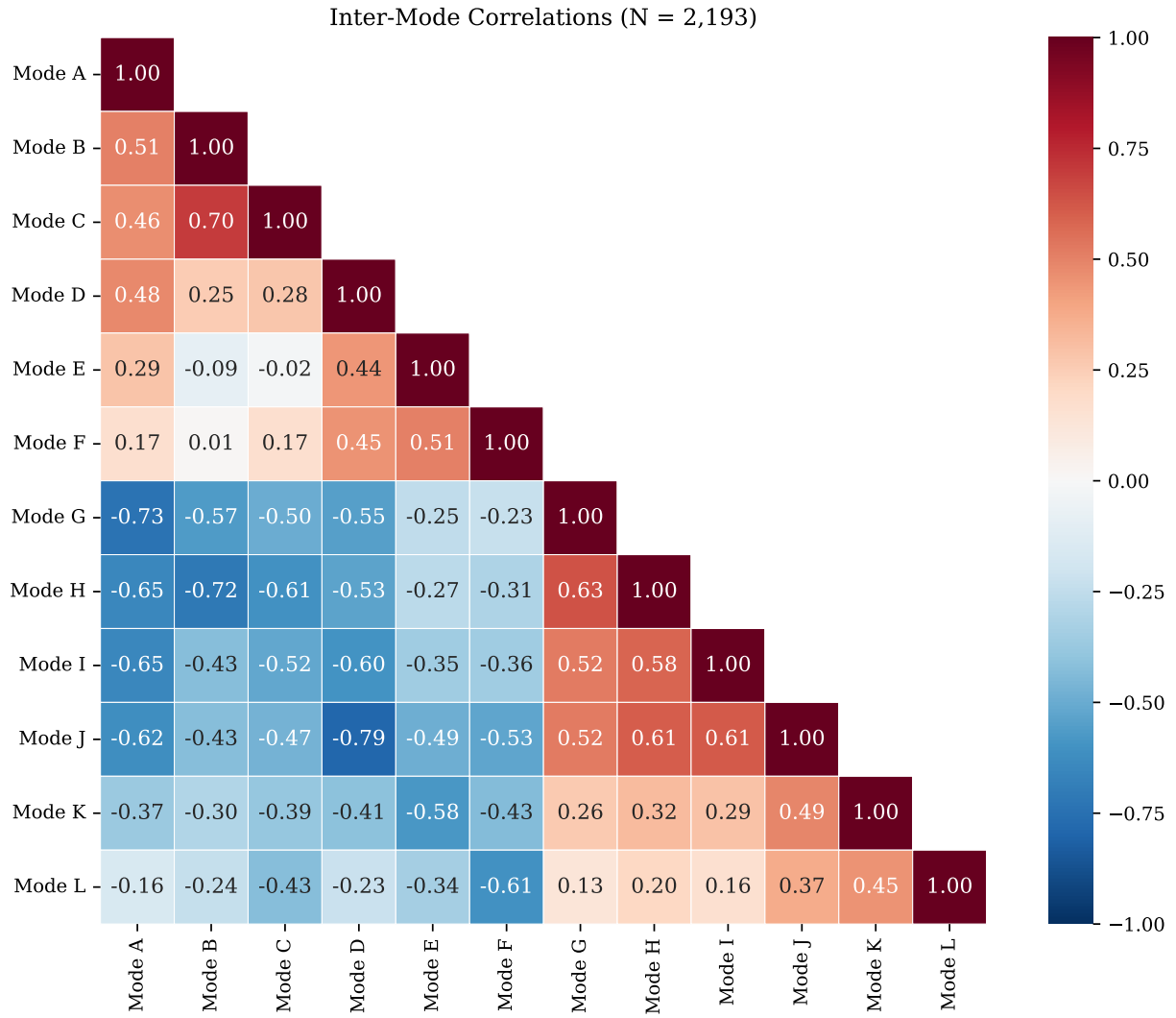


Figure 5: Inter-mode Pearson correlation matrix ($N = 2,193$). The block-diagonal structure of positive correlations within quadrants (A–C, D–F, G–H–I, J–K–L) and negative correlations between quadrants is consistent with the theoretical four-quadrant model. The pervasive negative correlations between Vision/Relation modes and Realization/Structuring modes reflect the forced-sum constraint inherent to ipsative data.

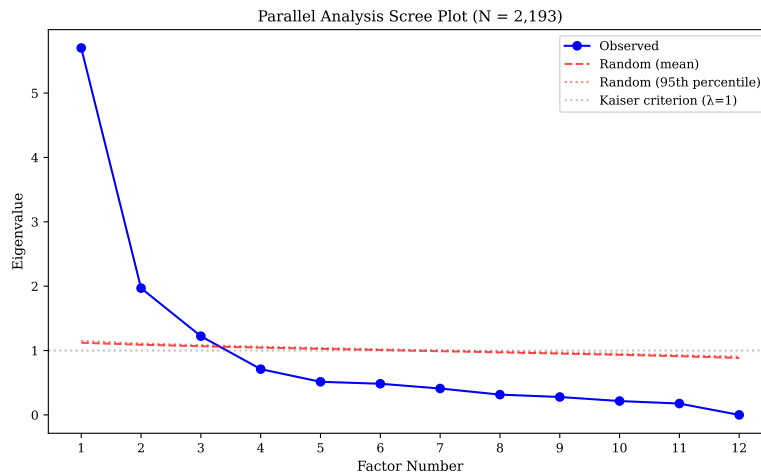


Figure 6: Scree plot with parallel analysis ($N = 2,193$). Black line: observed eigenvalues from PCA on ipsative scores. Dashed line: 95th percentile of eigenvalues from random data (parallel analysis). Three components clearly exceed the random threshold; the fourth (0.70) falls just below.

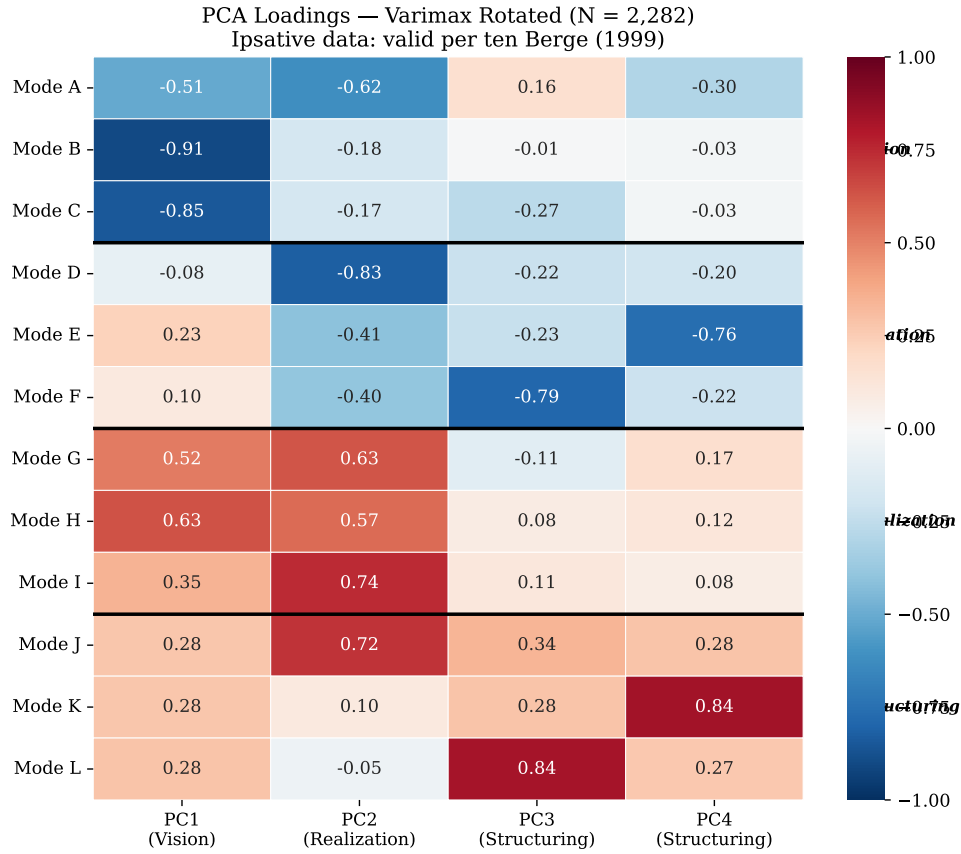


Figure 7: Varimax-rotated PCA loadings ($N = 2,193$, 4 components). Each row is a mode; each column is a component. The coherent block structure within quadrants supports the four-quadrant theoretical model.

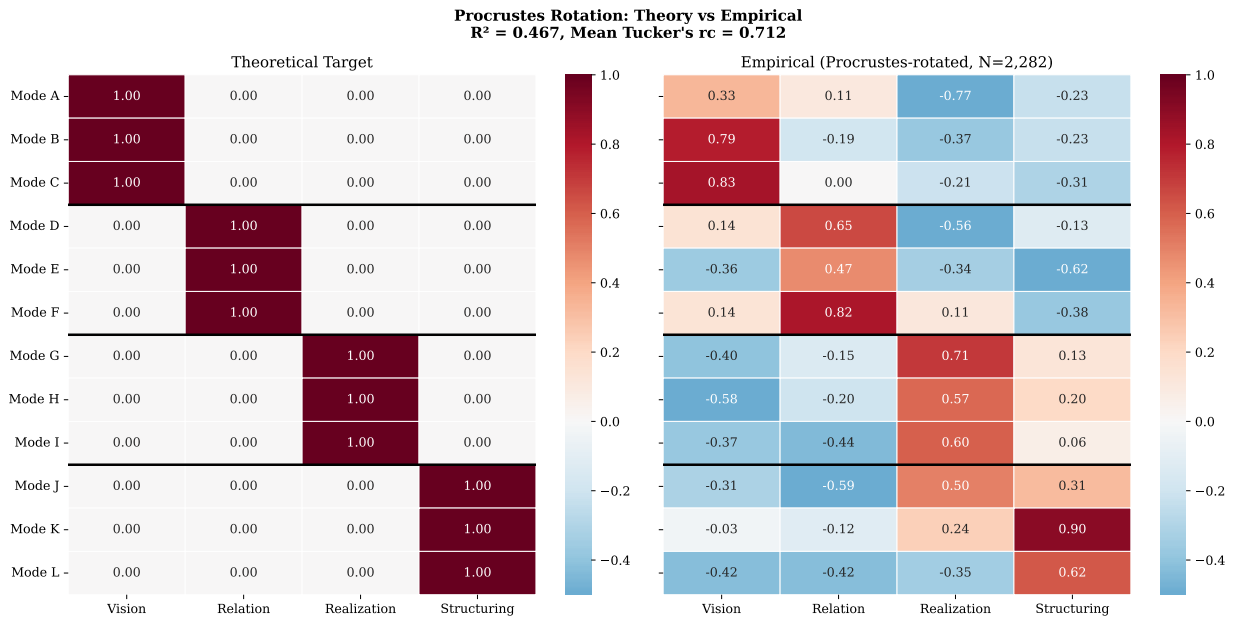


Figure 8: Procrustes target alignment of PCA-extracted components against the theoretical four-quadrant matrix. Tucker congruence coefficients per quadrant: Vision = .71, Relation = .76, Realization = .64, Structuring = .73 (mean $\phi = .71$).

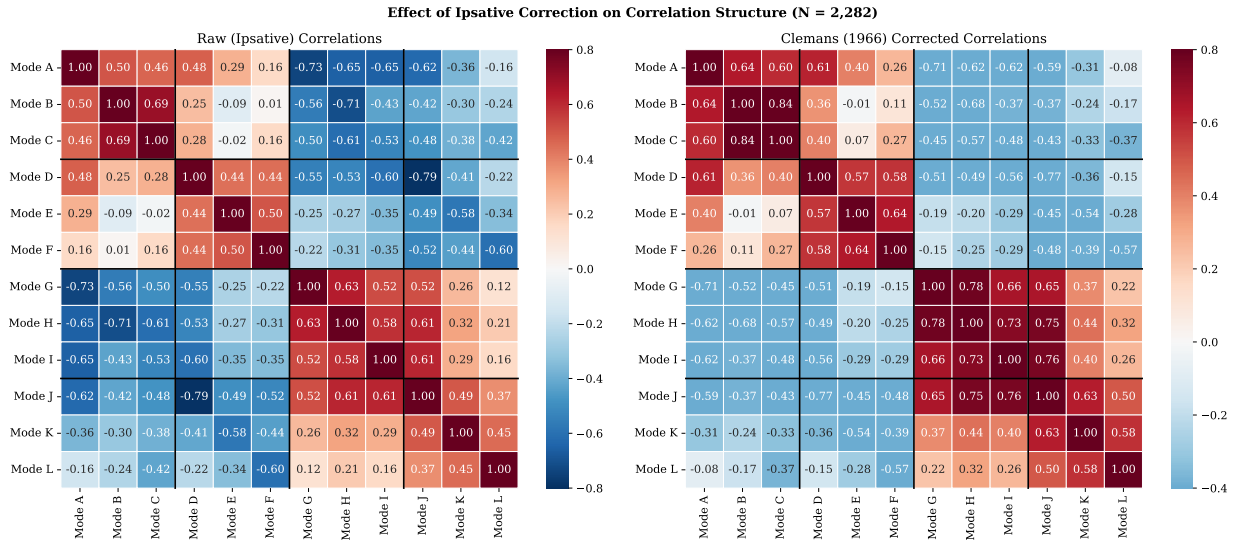


Figure 9: Inter-mode correlations before (left) and after (right) Clemans correction. The artificial negative bias from the fixed-sum constraint is removed, revealing the underlying block-diagonal structure consistent with the four-quadrant theoretical model.

4.3 Internal Consistency

Cronbach's alpha was computed within each mode on its four constituent descriptors. Results are presented in Table 7.

Table 7: Cronbach's α per Mode ($k = 4$ items each, $N = 2,282$)

Mode	α	95% CI low	95% CI high
A (Global Vision)	.665	.642	.687
B (Strategy)	.700	.679	.720
C (Dynamic)	.641	.616	.665
D (Self-Relationship)	.679	.657	.700
E (Understanding Others)	.690	.669	.710
F (Exchange)	.554	.524	.583
G (Execution)	.603	.576	.629
H (Organization)	.665	.642	.686
I (Management)	.579	.550	.606
J (Process)	.729	.711	.747
K (Understanding Ideas)	.424	.384	.461
L (Reference Framework)	.456	.419	.492
Mean	.615		

Ten of the 12 modes met the $\alpha \geq .50$ threshold, with a mean of $\alpha = .615$ across all modes. The two modes that fell below threshold are K (.424) and L (.456)—the most theoretical and abstract modes in the instrument, capturing conceptual analysis and reference-framework reasoning. The relatively restricted variance in these modes among executive samples—who as a group invest less energy in abstract intellectual content—contributes to the attenuated alpha, in line with the Spearman-Brown prediction for short scales.

4.4 Temporal Stability

Test-retest reliability was assessed via ICC(3,1) on $n = 85$ individuals who completed the instrument twice. Re-

sults are summarized in Table 8.

Table 8: Test-Retest ICC(3,1) per Mode and Quadrant ($n = 96$)

Mode / Quadrant	ICC	Pearson r
<i>Modes</i>		
A (Global Vision)	.737	.771
B (Strategy)	.723	.749
C (Dynamic)	.794	.806
D (Self-Relationship)	.656	.666
E (Understanding Others)	.726	.724
F (Exchange)	.608	.611
G (Execution)	.757	.780
H (Organization)	.715	.734
I (Management)	.641	.660
J (Process)	.679	.697
K (Understanding Ideas)	.724	.728
L (Reference Framework)	.610	.607
<i>Mean mode</i>	<i>.698</i>	<i>.711</i>
<i>Quadrants</i>		
Vision	.796	.831
Relation	.717	.718
Realization	.790	.818
Structuring	.708	.719
<i>Mean quadrant</i>	<i>.753</i>	<i>.772</i>

Applying Cicchetti's (1994) benchmarks (excellent: $\text{ICC} \geq .75$; good: .60–.74; moderate: .40–.59), two modes achieved “excellent” stability: C (.794) and G (.757). The remaining ten modes all fell in the “good” range (.60–.74): A (.737), B (.723), D (.656), E (.726), F (.608), H (.715), I (.641), J (.679), K (.724), L (.610). No mode dropped into the “moderate” band ($< .60$). At the quadrant level, Vision (.796) and Realization (.790) approached the “excellent” threshold, while Relation (.717) and Structuring (.708) sat comfortably in the “good” range. The mean mode-level ICC was .698 and the mean quadrant-level ICC was .753. The substantial

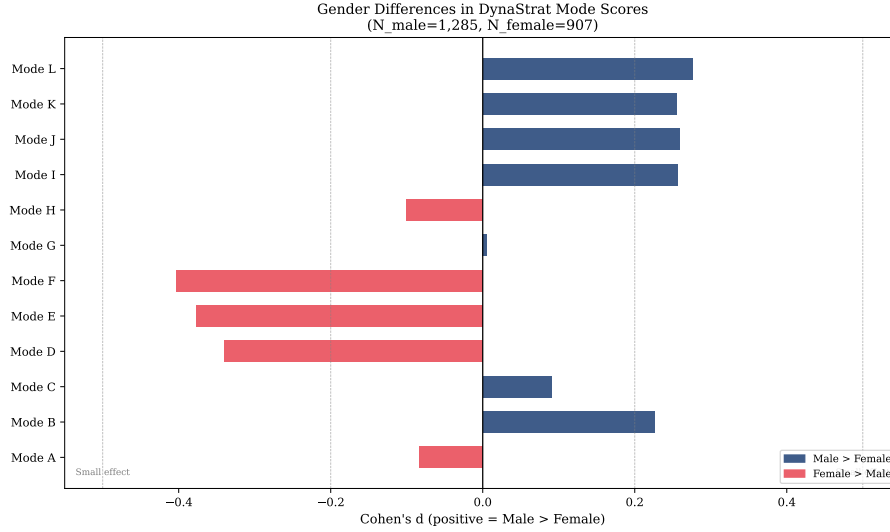


Figure 10: Gender effects (Cohen’s d , Male – Female) across the 12 DynaStrat modes ($N = 2,192$ with known gender). All effects fall within Cohen’s “small” range ($|d| < 0.50$).

heterogeneity in retest intervals (typical range: 41 days to ~ 3 years) is a limitation; the reported ICCs should therefore be interpreted as average stability across a heterogeneous range of intervals rather than as estimates at any specific time point.

4.5 Measurement Invariance Across Gender

Approximate multi-group CFA was conducted comparing male ($n = 1,285$) and female ($n = 907$) subsamples. Configural invariance was not established, as neither group achieved $\text{CFI} \geq .90$ in isolation—consistent with the artifactually poor fit expected for ipsative data. The comparison between groups yielded $\Delta\text{CFI} = .073$ and $\Delta\text{RMSEA} = -.167$, both exceeding the criteria of [Cheung & Rensvold \(2002\)](#) and [Chen \(2007\)](#). However, as noted by these authors, such criteria assume adequate baseline model fit; when applied to ipsative data where absolute fit is uniformly poor due to the singular covariance matrix, Δ indices may be unreliable indicators of invariance.

Given this limitation of the CFA-based approach, we supplemented the analysis with direct comparison of gender effects (Figure 10). Cohen’s d ranged from -0.40 to $+0.28$ across all 12 modes, with a mean $|d| = 0.22$. The largest effects were observed for Mode F ($d = -0.40$, favoring women) and Mode E ($d = -0.38$, favoring women), both in the Relation quadrant—consistent with established gender differences in interpersonal orientation. Men scored slightly higher on Realization mode I ($d = .26$) and Structuring quadrant modes L ($d = .28$), J ($d = .26$), and K ($d = .26$). All effects fell within Cohen’s “small” range ($|d| < 0.50$). These effect sizes are comparable to gender differences typically reported for personality instruments.

4.5.1 Gender $\times$ Age Interaction

To assess whether gender differences are stable across career stages or reflect cohort or maturational effects, we fit a moderated multiple regression for each of the 12 modes on the V1 analytic sample ($N = 2,192$; $n_{\text{male}} = 1,285$, $n_{\text{female}} = 907$; age $M = 46.5$, $SD = 9.5$, range 20–82). Table 9 reports, per mode, the gender main effect (β_{gender} , 95% CI, Holm-adjusted p , Bayesian BF_{10}) and the gender $\times$ age interaction (β_{int} , p , ΔR^2 , BF_{01} from BIC approximation). For descriptive comparison with the binned literature we additionally report quadrant-level two-way ANOVAs in Table 10 as supplementary.

Gender main effects. The Bayesian t -tests separated the 12 modes into a clear bimodal pattern. Eight modes (B, D, E, F, I, J, K, L) showed extreme support for a gender difference (BF_{10} from 3.8×10^4 to $> 10^{17}$, all Holm-adjusted $p < .001$): women allocated substantially more energy to the Relation modes D, E, F ($\beta \approx -7$ score points), while men allocated more to Structuring modes J, K, L and to Strategy (B) and Management (I) (β from $+4$ to $+5.8$). The remaining four modes (A, C, G, H) yielded anecdotal or sub-anecdotal evidence (BF_{10} from 0.05 to 0.69, all Holm-adjusted $p > .08$), consistent with negligible gender differences on Vision, Dynamic, Execution, and Organization.

Gender $\times$ Age interaction. Critically, the interaction was negligible on every mode. All 12 interaction coefficients fell within a narrow band ($|\beta_{\text{int}}| \leq 0.16$ score points per year, median 0.08), all Holm-adjusted $p > .98$, all $\Delta R^2 \leq .002$, and all Cohen’s $f^2 < 0.002$ —well below the conventional small-effect threshold of $f^2 = .02$ ([Cohen, 1988](#)). The Bayesian evidence converged on the same conclusion: BF_{01} ranged from 10.3 (Mode D) to 46.6 (Mode A), with a median of 31.4 across the 12 modes. By Jeffreys’ conventions, every

Table 9: Moderated Multiple Regression for Gender $\times$ Age on the 12 Mode Scores

Mode	Gender main effect				Gender $\times$ Age interaction			
	β_{gender}	95% CI	p_{Holm}	BF ₁₀	β_{int}	p_{Holm}	ΔR^2	BF ₀₁
A	-2.22	[-4.02, -0.43]	.108	0.31	+0.010	1.00	<.001	46.6
B	+4.48	[+2.68, +6.27]	<.001	3.8×10^4	+0.047	1.00	<.001	41.6
C	+1.39	[-0.31, +3.10]	.108	0.43	-0.122	1.00	.001	19.2
D	-6.92	[-8.59, -5.26]	<.001	$> 10^{12}$	+0.155	0.99	.001	10.3
E	-7.13	[-8.75, -5.51]	<.001	$> 10^{14}$	+0.135	1.00	.001	14.0
F	-6.97	[-8.42, -5.52]	<.001	$> 10^{17}$	+0.093	1.00	.001	23.0
G	+0.38	[-1.29, +2.04]	.892	0.05	-0.082	1.00	<.001	30.7
H	-1.52	[-3.15, +0.10]	.083	0.69	-0.036	1.00	<.001	43.0
I	+4.71	[+3.24, +6.19]	<.001	2.7×10^6	-0.014	1.00	<.001	46.1
J	+5.76	[+4.04, +7.48]	<.001	3.7×10^6	-0.080	1.00	<.001	32.2
K	+3.99	[+2.66, +5.33]	<.001	1.6×10^6	-0.096	1.00	.001	19.4
L	+4.07	[+2.84, +5.29]	<.001	7.9×10^7	-0.011	1.00	<.001	46.2

Note. $N = 2,192$. Gender is effects-coded ($M = +0.5$, $F = -0.5$); age is mean-centred. β_{gender} is the M–F mean difference at the sample mean age (in mode-score points). β_{int} is the change in the gender gap per additional year of age. p_{Holm} : Holm-adjusted p -value across the 12 tests in the family. BF₁₀ (gender): Bayesian independent-samples t -test with Cauchy prior $r = \sqrt{2}/2$. BF₀₁ (interaction): BIC-approximation Bayes Factor in favour of the no-interaction model.

Table 10: Supplementary Two-Way ANOVA: Gender $\times$ Age Band for Quadrant Scores

Variable	Gender			Age Band			Gender $\times$ Age		
	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Vision (A+B+C)	4.54	.033	.002	8.81	<.001	.016	0.38	.827	.001
Relation (D+E+F)	116.89	<.001	.051	1.54	.188	.003	1.17	.320	.002
Realization (G+H+I)	1.64	.200	.001	6.45	<.001	.012	0.41	.805	.001
Structuring (J+K+L)	59.16	<.001	.026	4.97	.001	.009	0.61	.658	.001

Note. $N = 2,192$. Age bands: < 30 ($n = 84$), $30\text{--}39$ ($n = 452$), $40\text{--}49$ ($n = 753$), $50\text{--}59$ ($n = 731$), ≥ 60 ($n = 172$). Reported for descriptive continuity only; primary inference is the MMR in Table 9.

mode reached the strong-evidence threshold ($\text{BF}_{01} \geq 10$), and the median crossed the very-strong threshold (≥ 30). We note, however, that Mode D sits at the lower boundary of the “strong” band ($\text{BF}_{01} = 10.3$, only marginally above the 10 cutoff), in part because $|\beta_{\text{int}}|$ for Mode D is the largest of the 12 (+0.155 score points per year). For this single mode the evidence is therefore best characterised as “strong but not unequivocal,” and a moderately less informative prior (e.g. Cauchy $r = 0.5$) could in principle move it into the moderate band. The remaining 11 modes carry comfortable margins above the strong threshold and would survive such reprior. In plain terms, for each mode the data are between 10 and 47 times more likely under a model in which the gender gap is constant across age than under a model in which it varies.

Supplementary: binned two-way ANOVAs. For continuity with the binned literature, Table 10 reports two-way ANOVAs (Gender $\times$ Age Band) at the quadrant level. The pattern reproduces the MMR findings: significant gender main effects on Relation ($p < .001$) and Structuring ($p < .001$), a significant age main effect on three of four quadrants, and uniformly non-significant interactions (all $p \geq .32$, all $\eta_p^2 \leq .002$). We caution that, by themselves, these non-significant interactions only license the weak claim “we failed to detect a Gender $\times$ Age effect”; the BF_{01} analysis above licenses

the substantively stronger “the data favour invariance.”

Figure 11 illustrates the invariance pattern at the quadrant level. The two lines (male/female) run approximately parallel across all five age bands for every quadrant: women consistently allocate more energy to Relation and men to Structuring, with the gap neither widening nor narrowing with age. Vision and Realization show minimal gender differences at every age.

To visualize the pattern at finer resolution, Figure 12 presents Cohen’s d for all 12 modes within each age band. Several observations emerge. First, the Relation modes (D, E, F) consistently favored women across all age bands, with effect sizes ranging from small to medium (Mode E: $d = -0.67$ in the < 30 band, -0.34 to -0.45 in older bands). Second, the Structuring modes (J, K, L) and Management mode (I) consistently favored men, with small effects stable across age ($d \approx +0.10$ to $+0.39$). Third, Mode C (Dynamic) shows the largest visual age-related drift in the binned heatmap, with a male advantage of $d = +0.49$ in the < 30 group attenuating to near zero in older cohorts. We flag this only as a descriptive observation: in the continuous-age MMR, the Mode C interaction yielded $\beta_{\text{int}} = -0.122$ score points per year ($p_{\text{Holm}} = 1.00$, $\Delta R^2 < .001$, $\text{BF}_{01} = 19.2$), so the appearance of an age-trend on Mode C is not distinguishable from sampling noise concentrated in the small < 30 band ($n = 84$). Readers should not interpret this visual feature as evi-

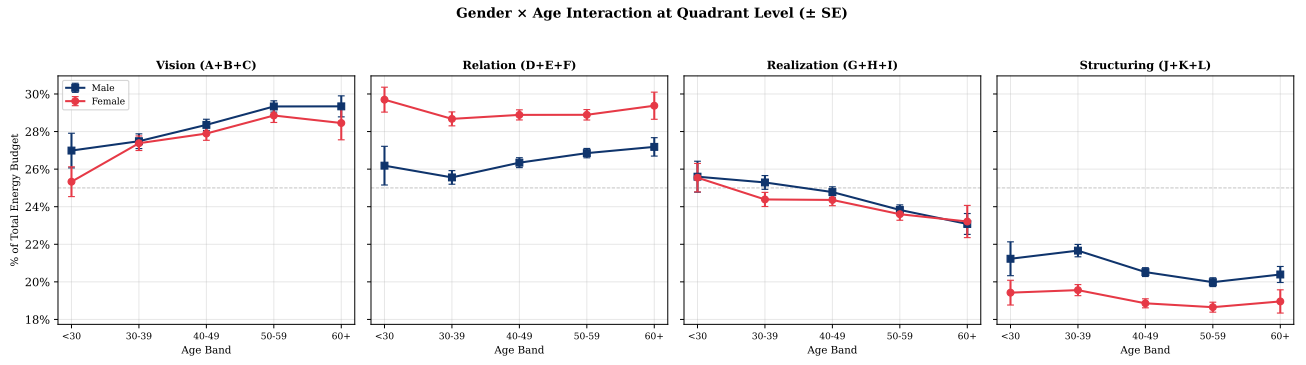


Figure 11: Gender $\times$ Age interaction at the quadrant level ($\pm$ SE). Lines for males and females run approximately parallel across all age bands, consistent with the strong Bayesian evidence for invariance (all $BF_{01} \geq 10$ at the mode level). The Relation quadrant shows the largest gender gap (favoring women), stable from the youngest to the oldest cohort. $N = 2,192$ (V1 only).

dence of a real age $\times$ gender effect.

As a descriptive supplement to the MMR, we computed the Pearson correlation between age band midpoint and within-band Cohen’s d for each mode. With only $k = 5$ data points per correlation, this test has very limited statistical power and should be interpreted as a trend summary rather than a formal hypothesis test (the continuous-age moderation in the MMR is the appropriate inferential target). None of the 12 modes showed a statistically significant linear trend (all $p > .12$; largest $|r| = .77$ for Mode I, $p = .13$). The mean absolute Cohen’s d across all modes was .35 in the < 30 band, .27 in 30–39, .22 in 40–49, .20 in 50–59, and .25 in the ≥ 60 band.

In summary, both gender and age exert independent main effects on DynaStrat mode scores, but they do not interact: the continuous-age moderation analysis returned negligible interaction coefficients (all $|\beta_{\text{int}}| \leq 0.16$ score points per year, all Holm-adjusted $p > .98$, all $\Delta R^2 \leq .002$), and Bayes Factors uniformly favoured the no-interaction model on every mode (median $BF_{01} = 31.4$, all ≥ 10). The data thus license a positive claim of invariance: gender differences in behavioral energy allocation are stable across the age range represented in this executive sample (ages 20–82), not merely failing to register as significant.

5 Discussion

This study provides the first comprehensive psychometric examination of DynaStrat, an ipsative energy-allocation instrument used in executive assessment since 2011. Using archival data from 2,193 executive profiles collected between 2020 and 2026, we employed multiple analytic approaches adapted for ipsative data to evaluate structural validity, reliability, temporal stability, and measurement invariance.

5.1 Summary of Findings

Formal CFA-based measurement invariance across gender was not established, a finding attributable in part to the artifactually poor baseline model fit inherent to

ipsative data. Direct comparison of mean scores revealed uniformly small gender effects (mean $|d| = 0.22$), though small mean differences do not substitute for formal invariance testing. Whether the instrument functions equivalently across genders at the measurement level remains an open question. However, the moderated multiple regression with continuous age provided additional indirect evidence for the structural nature of these differences: all 12 interaction coefficients were negligible ($|\beta_{\text{int}}| \leq 0.16$ score points per year), and the Bayes Factor analysis returned strong-to-very-strong evidence for the no-interaction model on every mode (median $BF_{01} = 31.4$, all ≥ 10). The fact that the data positively favour invariance—rather than merely failing to reject it—is inconsistent with the hypothesis that observed gender differences are artifacts of cohort effects, differential career socialization, or age-related changes in self-presentation, and instead suggests that they reflect genuine differences in how men and women allocate behavioral energy—a finding consistent with the broader personality literature (Costa et al., 2001; Schmitt et al., 2008).

5.2 Implications for Ipsative Measurement

Our findings contribute to the ongoing debate about the psychometric viability of ipsative instruments (Baron, 1996; Brown & Maydeu-Olivares, 2011). The pattern of results—good-to-excellent stability, adequate internal consistency, and recoverable structure despite artifactually poor classical fit indices—is precisely what the theoretical literature predicts for well-constructed forced-choice instruments analyzed with appropriate methods.

The ipsative constraint represents both the instrument’s greatest strength (ecological validity, resistance to response biases) and its primary psychometric challenge (non-independent scores, singular covariance matrices). We addressed this tension through methodological pluralism: rather than relying on a single analytic approach, we employed multiple complementary methods (Clemans correction, drop-one CFA, PCA recovery, Procrustes rotation) that converge on the same struc-

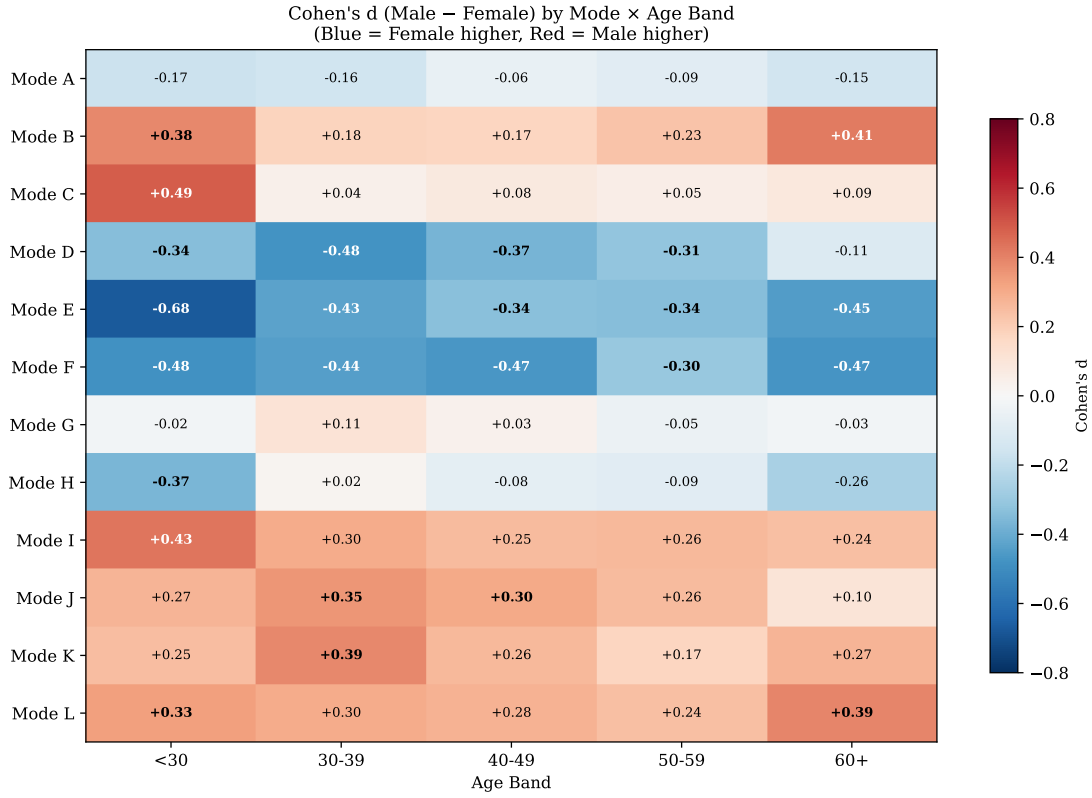


Figure 12: Cohen's d (Male – Female) by mode and age band. Red cells indicate male-favoring effects; blue cells indicate female-favoring effects. The consistency of color within each row (mode) illustrates the stability of gender differences across age. $N = 2,192$ (V1 only).

tural conclusion. This multi-method approach may serve as a template for validating other ipsative instruments.

5.3 Findings in the Context of Established Instruments

Two aspects of our results are best understood by reference to the broader literature on normative instruments such as the Big Five and MBTI, which we contrasted on design grounds in Section 1.1.

First, the *magnitude* of gender effects observed here (mean $|d| = 0.22$, largest $|d| = 0.40$ on Mode F) is broadly comparable to what is typically reported for normative personality instruments. The most-cited cross-cultural Big Five investigation (Costa et al., 2001) reports gender effects on Neuroticism around $d \approx +0.25$ and on Agreeableness around $d \approx +0.15$ favouring women, with men slightly higher on Openness-to-Ideas; the larger 55-culture replication by Schmitt et al. (2008) reports broader ranges (Neuroticism d up to $\approx +0.40$ in some samples). Our observed effects sit squarely within this normative range, suggesting that the ipsative scoring metric does not in itself dramatically compress gender differences in this domain. A theoretical caveat applies, however: under a fixed-sum constraint, a between-group difference on one mode is mechanically offset by opposite-direction differences on other modes, so the *distribution* of differences across modes is constrained even when group means genuinely

differ. The interpretation that follows is not about absolute trait levels (where men and women could in principle differ uniformly) but about *relative allocations* (where one group's higher allocation on Relation must be offset somewhere else—in this sample, on Structuring and Management). Practitioners interpreting DynaStrat gender effects should therefore read them as relative emphasis differences rather than as additive trait gaps.

Second, the *pattern* of gender effects is qualitatively congruent with the normative literature. Women allocated more energy to the three Relation modes (D, E, F: Self-Relationship, Understanding Others, Exchange), broadly consistent with the established findings of higher Agreeableness, higher communal orientation, and higher interpersonal sensitivity in women on normative measures (Costa et al., 2001; Schmitt et al., 2008). Men allocated more energy to the three Structuring modes (J, K, L: Process, Understanding Ideas, Reference Framework) and to Management (I), reflecting a pattern that parallels normatively-measured tendencies toward analytical and systemizing content. The convergence on this qualitative pattern—despite the very different scoring metric—provides indirect cross-instrument validation. Establishing this convergence formally through direct convergent-validity studies against the NEO-PI-R or HEXACO remains an important gap, listed in our future directions below.

Third, the *stability* of gender effects across age (median $BF_{01} = 31.4$, all 12 modes ≥ 10) is also consistent

with the Big Five literature, which has generally found gender differences to be remarkably stable across the adult lifespan once measurement is appropriately modelled (Costa et al., 2001; Schmitt et al., 2008). The Bayesian framework adopted here strengthens this convergence: rather than failing to detect a developmental change (which would leave the question open), we report positive evidence that the gender gap is constant across age—a stronger and more interpretable result than the null findings typical of binned ANOVA designs in the comparison literature.

5.4 Limitations and Future Directions

Several limitations should be acknowledged. First, the sample is predominantly French executives, limiting the generalizability of the normative data to other cultural contexts. The assessment was available in seven languages, but the test language field was not consistently recorded in the archival database, precluding language-based invariance analyses. Second, the archival nature of the data means that external criterion variables (e.g., leadership performance ratings, 360-degree assessments) were not available for criterion-related validity analyses. Third, Modes K and L showed below-threshold internal consistency. As discussed, these are the most theoretical modes, and the executive sample invests comparatively little energy in them, producing restricted variance that attenuates alpha; nonetheless, their constituent descriptors may also benefit from revision. Fourth, convergent validity with established normative instruments (e.g., NEO-PI-R, MBTI) was not assessed in the present study and remains an important gap; such analyses are needed to establish construct validity. Fifth, while the Clemans correction and drop-one approaches improved structural fit, a full Thurstonian IRT analysis (Brown & Maydeu-Olivares, 2011, 2012) using item-level forced-choice responses would provide the most rigorous test of latent structure and is recommended for future research.

Future studies should examine: (a) criterion-related validity against external measures of leadership effectiveness; (b) convergent validity with established normative personality instruments (e.g., NEO-PI-R, MBTI); (c) cross-cultural invariance using language-tagged data; (d) Thurstonian IRT modeling at the item level; (e) longitudinal studies examining profile changes in response to coaching interventions; (f) revision of the weaker modes (K, L) to improve internal consistency; and (g) psychometric validation of the revised V2 instrument once sufficient data are available.

5.5 Practical Implications

For practitioners, the results support the use of DynaStrat in executive coaching and leadership development contexts. The good-to-excellent temporal stability indicates that profiles can be meaningfully discussed over extended coaching engagements without concern that observed patterns are merely transient fluctuations. Formal measurement invariance across gender was not

established, and mean-level gender differences—while uniformly small (mean $|d| = 0.22$, largest $|d| = 0.40$)—do not by themselves guarantee equivalent measurement. Pending formal invariance testing with appropriate methods (e.g., Thurstonian IRT), practitioners should interpret gender-related score differences with caution. However, the demonstrated stability of gender effects across the full age range—supported by strong Bayesian evidence on every mode (all $BF_{01} \geq 10$ for the gender $\times$ age interaction, continuous-age MMR)—provides positive evidence that these differences are not artifacts of career stage and do not require age-specific correction. The magnitude of observed differences does not suggest a need for separate gender norms at this time. The quadrant-level aggregation provides the most reliable measurement unit ($ICC \geq .70$) and should be emphasized in feedback sessions, while mode-level interpretations should be offered as more granular but less stable insights.

6 Conclusion

DynaStrat demonstrates psychometric properties consistent with a well-constructed ipsative instrument: good-to-excellent temporal stability (mean quadrant $ICC = .753$, mean mode $ICC = .698$), adequate internal consistency for 10 of 12 modes (mean $\alpha = .615$), a four-quadrant structure recoverable through ipsative-appropriate methods, and gender differences that are small in magnitude and shown to be invariant across age by continuous-moderator MMR and Bayesian model comparison (all 12 modes: $BF_{01} \geq 10$ in favour of the no-interaction model). These strengths must be weighed against notable limitations: no convergent evidence against established external instruments (e.g., NEO-PI-R), no criterion-related validity data, two modes (K and L) with below-threshold internal consistency, and structural fit indices that remain below conventional thresholds even after ipsative corrections. The forced-distribution design, while creating these analytic challenges, ensures ecological validity and resistance to response biases that are particularly valuable in high-stakes organizational assessment. These findings establish an initial psychometric foundation for DynaStrat’s use in executive coaching and leadership development, while highlighting specific areas—external validation, item revision, and Thurstonian IRT modeling—as priorities for future research.

Author Contributions

Patrick Dura: Conceptualization, Instrument Development, Data Curation, Formal Analysis, Writing—Original Draft, Project Administration. Daniel Bos: Conceptualization, Methodology, Validation, Writing—Review & Editing. Clément Poirer: Methodology, Statistical Analysis, Validation, Writing—Review & Editing. Marion Noulhiane: Methodology, Validation, Writing—Review & Editing, Supervision.

Conflict of Interest Statement

The authors are affiliated with Stratelio, the organization that developed and distributes the DynaStrat instrument. This relationship is disclosed in the interest of full transparency. The psychometric analyses reported here were conducted using standard methods and publicly available statistical procedures; all analytic code and summary tables are available from the corresponding author upon request.

Ethics Statement

All data were archival, collected in the normal course of professional practice between 2020 and 2026. All participants provided informed consent for the use of their anonymized data for research purposes as part of the standard assessment agreement. Data were fully anonymized prior to analysis. Under French law (Loi Jardé, Article R1121-1 of the Code de la Santé Publique), retrospective analysis of anonymized non-interventional professional data does not constitute research involving the human person and is therefore exempt from ethics committee (CPP) review. Data processing complied with the European General Data Protection Regulation (GDPR, Regulation 2016/679).

Funding

This research received no external funding. All work was conducted as part of Stratelio's ongoing instrument development and quality assurance activities.

Data Availability Statement

The data that support the findings of this study cannot be shared publicly due to confidentiality agreements with the organizations that commissioned the assessments. Anonymized summary statistics and correlation matrices are available from the corresponding author upon reasonable request.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.
- Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, 69(1), 49–56. doi:10.1111/j.2044-8325.1996.tb00599.x
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54(1), 106–148. doi:10.1111/j.1467-6494.1986.tb00391.x
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460–502. doi:10.1177/0013164410375112
- Brown, A., & Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods*, 44(4), 1135–1147. doi:10.3758/s13428-012-0217-x
- Chan, W., & Bentler, P. M. (1993). The covariance structure analysis of ipsative data. *Sociological Methods & Research*, 22(2), 214–247. doi:10.1177/0049124193022002005
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. doi:10.1080/10705510701301834
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255. doi:10.1207/S15328007SEM0902_5
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. doi:10.1037/1040-3590.6.4.284
- Clemans, W. V. (1966). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs*, 14.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81(2), 322–331. doi:10.1037/0022-3514.81.2.322
- Dunlap, W. P., & Cornwell, J. M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research*, 29(1), 115–126. doi:10.1207/s15327906mbr2901_5

- Goldberg, L. R. (1990). An alternative “description of personality”: The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216–1229. doi:10.1037/0022-3514.59.6.1216
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74(3), 167–184. doi:10.1037/h0029780
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. doi:10.1080/10705519909540118
- Inscape Publishing / Wiley. (2013). *Everything DiSC research report*. Wiley.
- Jung, C. G. (1921/1971). *Psychological types* (H. G. Baynes, Trans., revised by R. F. C. Hull). Princeton University Press.
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Li, M., Zhang, B., & Mou, Y. (2025). Though forced, still valid: Examining the psychometric performance of forced-choice measurement. *Psychological Assessment*.
- Loewenthal, K. M. (2001). *An introduction to psychological tests and scales* (2nd ed.). Psychology Press.
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker’s congruence coefficient as a meaningful index of factor similarity. *Methodology*, 2(2), 57–64. doi:10.1027/1614-2241.2.2.57
- Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, 77(4), 531–551. doi:10.1348/0963179042596450
- Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Consulting Psychologists Press.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212. doi:10.1016/j.jrp.2006.02.001
- Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can’t a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, 94(1), 168–182. doi:10.1037/0022-3514.94.1.168
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. doi:10.1037/0033-2909.86.2.420
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804. doi:10.3758/BF03194105
- ten Berge, J. M. F. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research*, 34(1), 89–102. doi:10.1207/s15327906mbr3401_4