Theory: Formal Analysis Tools and Interpretive Meaning
Profile Pattern Diagnostics (PPD)

Pai Surya Darshan

June 2021 (Refined in Jan 2026)

Abstract

This document describes the formal analytical tools used in **Profile Pattern Diagnostics (PPD)** and their intended interpretive meaning. The system is **explicitly** diagnostic: all constructs are designed to describe structure in rubric-based skill profiles without producing rankings (until explicitly requested beyond the scope of this code), predictions, or guidance.

1. Input normalisation

1.1. Linear normalisation to $[0, 1]$

Formal definition

$$x_{\text{norm}} = \frac{x}{5} \tag{1}$$

where,

- x is the original rubric score on a 0–5 scale
- $x_{\text{norm}} \in [0, 1]$ is the normalised score

Verbal meaning

Maps rubric-based scores to a unitless interval while preserving relative magnitude and ordinal relationships between dimensions. The transformation does not alter the internal structure of the profile, only its numerical scale.

Reasoning

Ensures consistent scaling across all dimensions and groups, enabling shared thresholds and descriptors without embedding rubric-specific constants into downstream logic. This avoids population dependence and preserves interpretability across datasets.

1.2. Justification for choice of Kendall-aligned thresholds

Thresholds used throughout PPD are aligned with Kendall-style ordinal interpretation rather than distribution-dependent calibration. Because rubric scores are inherently ordinal and assessor-driven, thresholds are chosen to respect relative ordering and monotonic relationships rather than assuming interval-level precision or population normality.

Reasoning

Kendall-aligned thresholds are robust to assessor bias, scale compression, and cohort composition. They preserve the rank-consistent meaning of scores under linear normalisation, avoid dependence on population statistics, and remain stable across cohorts of different sizes and skill distributions. This choice ensures that all gates and pattern checks operate on *structural order* rather than inferred quantitative distance, maintaining diagnostic validity under minimal assumptions.

2. Summary descriptors (order-invariant, within-candidate)

2.1. Mean (μ)

Formal definition

$$\mu = \frac{1}{n}\sum_{i=1}^{n} r_i \tag{2}$$

where,

- r_i is the score of the i-th dimension
- n is the total number of dimensions

Verbal meaning

Represents the overall level of capability within the profile, summarising the general magnitude of performance across all dimensions without regard to internal distribution.

Reasoning

Used to determine level-based tiers only after balance has been established. The mean is treated as meaningful only when the profile exhibits internal coherence; otherwise, it may mask structural unevenness.

2.2. Standard deviation (σ, population)

Formal definition

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \mu)^2} \tag{3}$$

where,

- r_i is the score of the i-th dimension
- μ is the mean score
- n is the number of dimensions

Verbal meaning

Quantifies the degree of dispersion across dimensions, indicating how evenly or unevenly capability is distributed within a single profile.

Reasoning

Acts as the balance gate for the profile. Population standard deviation is used because the dimensions form a complete, fixed set rather than a sample drawn from a larger population.

2.3. Minimum and bottleneck dimension

Formal definition

$$m = \min_i r_i, \qquad d_{\min} = \arg\min_i r_i \tag{4}$$

where,

- r_i is the score of the i-th dimension
- m is the minimum observed score

- $d_{\min}$ identifies the weakest dimension

Verbal meaning

Identifies the weakest dimension in the profile and quantifies its magnitude, highlighting the point at which performance is most constrained.

Reasoning

Encodes a failure-first structural constraint. In many applied contexts, overall effectiveness is limited by the weakest component rather than the average.

2.4. Maximum (contextual only)

Formal definition

$$M = \max_i r_i \tag{5}$$

where,

- r_i is the score of the i-th dimension
- M is the maximum observed score

Verbal meaning

Represents the strongest observed dimension within the profile.

Reasoning

Used solely for contextual interpretation and not as a gating or tiering criterion.

2.5. Range (Δ)

Formal definition

$$\Delta = \max(\mathbf{r}) - \min(\mathbf{r}) \tag{6}$$

where,

- $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ is the vector of dimension scores
- $\max(\mathbf{r})$ and $\min(\mathbf{r})$ denote the maximum and minimum elements of the vector

Verbal meaning

Measures the contrast between the strongest and weakest dimensions.

Reasoning

Complements dispersion-based descriptors by capturing unevenness intensity in a single quantity.

3. Order-aware descriptor (within-candidate)

3.1. Adjacent difference energy (D)

Formal definition

$$D = \frac{1}{n} \sum_{i=1}^{n} |r_i - r_{i+1}|, \qquad r_{n+1} = r_1 \tag{7}$$

where,

- r_i is the score of the i-th ordered dimension
- n is the total number of dimensions

- cyclic indexing enforces rotational invariance

Verbal meaning

Quantifies local volatility between neighbouring dimensions.

Reasoning

Applied only when axis ordering is semantically meaningful.

4. Pattern classification constructs (within-candidate)

4.1. Balanced gate

Formal rule

$$\text{balanced} \iff \sigma \leq \tau_{\text{bal}} \tag{8}$$

where,

- σ is the within-candidate standard deviation
- τ_{bal} is a predefined balance threshold

Verbal meaning

Indicates consistent development across all dimensions.

Reasoning

Prevents level-based interpretation of structurally uneven profiles.

4.2. Uniform patterns (level-conditional, balanced-only)

Verbal meaning

Uniform patterns are level descriptors applied only after internal balance has been established. They distinguish uniformly low and uniformly high profiles without introducing ranking, and without overriding structural diagnostics.

Reasoning

Uniformity is only meaningful when dispersion is already below the balance threshold. Applying level labels to uneven profiles would collapse structure into a single number and can mask bottlenecks or polarity. These checks therefore operate *within* the balanced regime and are interpreted as descriptive tiers only.

4.2.1. Uniform low (balanced and low level)

Formal rule

$$\text{uniform\_low} \iff \big(\sigma \leq \tau_{\text{bal}}\big) \ \wedge \ \big(\mu \leq \tau_{\text{low,uni}}\big) \tag{9}$$

where,

- σ is the within-candidate standard deviation
- μ is the within-candidate mean
- τ_{bal} is the predefined balance threshold
- $\tau_{\text{low,uni}}$ is a predefined uniform-low mean threshold

Verbal meaning

Indicates a consistently developed profile whose overall level is uniformly low.

Reasoning

Flags a coherent but low-development profile. This is descriptive only and does not imply suitability or ranking.

4.2.2. Uniform high (balanced and high level)

Formal rule

$$\text{uniform\_high} \iff \left(\sigma \le \tau_{\text{bal}}\right) \wedge \left(\mu \ge \tau_{\text{high,uni}}\right) \tag{10}$$

where,

- σ is the within-candidate standard deviation
- μ is the within-candidate mean
- τ_{bal} is the predefined balance threshold
- $\tau_{\text{high,uni}}$ is a predefined uniform-high mean threshold

Verbal meaning

Indicates a consistently developed profile whose overall level is uniformly high.

Reasoning

Flags a coherent and high-development profile. This is descriptive only and does not imply suitability or ranking.

4.2.3. Uniform mid (balanced ∧ mid level; optional)

Formal rule

$$\text{uniform\_mid} \iff \left(\sigma \le \tau_{\text{bal}}\right) \wedge \left(\tau_{\text{low,uni}} < \mu < \tau_{\text{high,uni}}\right) \tag{11}$$

where,

- σ is the within-candidate standard deviation
- μ is the within-candidate mean
- τ_{bal} is the predefined balance threshold
- $\tau_{\text{low,uni}}$ and $\tau_{\text{high,uni}}$ define the uniform-level band

Verbal meaning

Indicates a balanced profile whose overall level is neither uniformly low nor uniformly high.

Reasoning

Provided as an optional descriptive bucket to complete level partitioning within the balanced regime. It remains non-ranking and does not override other structural diagnostics.

4.3. Bottlenecked pattern

Formal rule

$$\min(\mathbf{r}) \le \tau_{\text{bottleneck}} \tag{12}$$

where,

- $\mathbf{r}$ is the vector of dimension scores
- $\tau_{\text{bottleneck}}$ is a predefined bottleneck threshold

Verbal meaning

Signals the presence of a structurally limiting weak dimension.

Reasoning

Ensures that only meaningful constraints are flagged.

4.4. Polarised pattern

Formal rule

$$\Big(\max(\mathbf{r}) \geq \tau_{\text{high}}\Big) \wedge \Big(\min(\mathbf{r}) \leq \tau_{\text{low}}\Big) \wedge \Big(\text{range}(\mathbf{r}) \geq \tau_{\text{pol}}\Big) \tag{13}$$

where,

- $\mathbf{r}$ is the vector of dimension scores
- τ_{high} is a predefined high-score threshold
- τ_{low} is a predefined low-score threshold
- τ_{pol} is a predefined polarisation (spread) threshold
- $\text{range}(\mathbf{r}) = \max(\mathbf{r}) - \min(\mathbf{r})$

Verbal meaning

Signals the coexistence of pronounced strengths and pronounced weaknesses within the same candidate profile.

Reasoning

The polarised gate establishes the *existence* of structural polarity using absolute thresholds on extrema and spread.

4.4.1. Standardised scores

Once polarity is established, individual dimension contributions may be quantified using within-candidate standardised scores (z-scores), defined as:

$$z_i = \frac{r_i - \mu}{\sigma} \tag{14}$$

where,

- r_i is the score of dimension i
- μ is the mean of $\mathbf{r}$
- σ is the within-candidate standard deviation

Positive z-scores indicate dimensions that are elevated relative to the candidate's own average, while negative z-scores indicate dimensions that are depressed. This separation enables clear identification of high-polarised and low-polarised dimensions without introducing additional decision thresholds. The use of z-scores preserves scale invariance and prevents attribution based solely on absolute score magnitude or ordinal ranking.

5. Across-candidate descriptors (within-group)

5.1. Cohort mean (μ_c)

Formal definition

$$\mu_c = \frac{1}{N}\sum_{j=1}^{N} x_j \tag{15}$$

where,

- x_j is the score of a fixed dimension for candidate j

- N is the total number of candidates in the cohort

Verbal meaning

Average level of a fixed dimension across a cohort.

Reasoning

Used for programme-level assessment, not individual evaluation.

5.2. Sample standard deviation across candidates (s_c)

Formal definition

$$s_c = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(x_j - \mu_c)^2} \tag{16}$$

where,

- x_j is the score of a fixed dimension for candidate j
- μ_c is the cohort mean for that dimension
- N is the number of candidates

Verbal meaning

Quantifies variability across candidates for a fixed dimension.

Reasoning

Appropriate when candidates are treated as samples from a larger population.

5.3. Threshold breach rate

Formal definition

$$p_{\text{breach}} = \frac{1}{N}\sum_{j=1}^{N}\mathbb{I}(x_j \leq \tau) \tag{17}$$

where,

- x_j is the score of a fixed dimension for candidate j
- τ is a predefined operational threshold
- $\mathbb{I}(\cdot)$ is the indicator function
- N is the number of candidates

Verbal meaning

Proportion of candidates falling below a defined operational threshold.

Reasoning

Provides a non-ranking, ethical measure of cohort-level risk.

5.4. Percentile rank (optional)

Formal definition

$$\text{percentile}(x_j) = \frac{\text{rank}(x_j) - 1}{N-1} \tag{18}$$

where,

- x_j is the score of candidate j

- $\mathrm{rank}(x_j)$ is the ordinal rank within the cohort
- N is the number of candidates

Verbal meaning

Relative position of a candidate within the cohort.

Reasoning

Provided strictly as optional, contextual information for cohort interpretation. Group-level percentile ranks are excluded from diagnostics, gating, tiering, and decision-making, and are intended solely to support descriptive cohort analysis and programme-level planning.

6. Across-candidate descriptors (across-group)

6.1. Group aggregate score

Formal definition

$$G_{k,j} = \frac{1}{n_k} \sum_{i=1}^{n_k} r_{k,i,j} \tag{19}$$

where,

- $r_{k,i,j}$ is the i-th dimension of group k for candidate j
- n_k is the number of dimensions in group k

Verbal meaning

Represents a candidate's overall level within a capability group.

Reasoning

Defines a transparent, non-weighted aggregation required for group-level cohort analysis.

6.2. Group-level cohort variability

Formal definition

$$s_{G_k} = \sqrt{\frac{1}{N-1} \sum_{j=1}^{N} (G_{k,j} - \mu_{G_k})^2} \tag{20}$$

where,

- $G_{k,j}$ is the group-level aggregate score for candidate j
- μ_{G_k} is the mean group-level score across candidates
- N is the number of candidates

Verbal meaning

Measures how much candidates differ from each other at the group level.

Reasoning

Supports programme prioritisation and training design without collapsing individual structure.

6.3. Group-level breach rate

Formal definition

$$p_{\text{breach},k} = \frac{1}{N}\sum_{j=1}^{N} \mathbb{I}(G_{k,j} \le \tau_k) \tag{21}$$

where,

- $G_{k,j}$ is the group-level aggregate score for candidate j
- τ_k is the threshold for capability group k
- $\mathbb{I}(\cdot)$ is the indicator function
- N is the number of candidates

Verbal meaning

Proportion of candidates whose group-level capability falls below a defined threshold.

Reasoning

Used exclusively for strategic planning and cohort-level intervention assessment.

6.4. Group-level percentile rank (optional)

Formal definition

$$\text{percentile}(G_{k,j}) = \frac{\text{rank}(G_{k,j}) - 1}{N - 1} \tag{22}$$

where,

- $G_{k,j}$ is the aggregate score of group k for candidate j
- $\text{rank}(G_{k,j})$ is the rank of candidate j within group k
- N is the number of candidates

Verbal meaning

Indicates a candidate's relative position within the cohort for a given capability group.

Reasoning

Provided strictly as optional, contextual information for cohort interpretation. Group-level percentile ranks are excluded from diagnostics, gating, tiering, and decision-making, and are intended solely to support descriptive cohort analysis and programme-level planning.

7. Explicit exclusions

The following are intentionally excluded:

- Standardisation (z-scores)
- Population-relative metrics
- Composite quality scores
- Ranking or suitability inference
- Personality-based analytics

These exclusions are design choices, not limitations.

8. Summary mental model

- Mean → level

- Standard deviation → balance
- Minimum → bottleneck
- Range → contrast
- Polarisation and z-score → relative extremity (within-candidate attribution)
- D → local volatility
- Pattern labels → structural descriptors

All outputs are descriptive diagnostics only.

Author's Note

Responsibility for all analytical tools, underlying theory, metric selection, and overall system design rests entirely with my independent decisions. The preparation of this document was assisted by ChatGPT (OpenAI) solely for drafting and formatting **of the back-end** LaTeXsupport.