CHS Threshold Adjustment Guide

Administrator Reference for Configuring Composite Health Scores

This guide helps administrators understand and configure the Composite Health Score (CHS) system without requiring deep statistical knowledge. Use this reference when evaluating whether to adjust default settings for your organization.

Part 1: Parameter Reference Table

Component Weights

Parameter	Default	Purpose	Range	When to Adjust
CSS weight	0.50 (50%)	How much current state matters	0.30-0.70	Increase if "where teams are now" matters more than their trajectory
TRS weight	0.35 (35%)	How much improvement trajectory matters	0.15-0.50	Increase if rewarding improvement is more important than current position
PGS weight	0.15 (15%)	How much peer comparison matters	0.00-0.30	Decrease for standalone evaluation; increase for competitive benchmarking

Statistical Parameters

Parameter	Default	Purpose	Range	When to Adjust
Shrinkage kappa	10	Controls how much small samples pull toward average	5-20	Increase for more conservative estimates with small groups
Average correlation	0.30	Assumed correlation between indicators	0.10-0.50	Adjust only if you have empirical evidence about indicator independence
SE inflation	1.20	Accounts for component covariance	1.10-1.40	Keep default unless validated with your data
Z-score winsorization	3.0	Caps extreme indicator values	2.5-3.5	Lower to be more conservative about outliers
TRS winsorization	4.5	Caps extreme trajectory values	3.5-5.0	Lower if seeing unrealistic trajectory scores

Category Thresholds

Category	Default Threshold	Score Range	When to Adjust
Excellent	70	70-100	Lower to 65 if too few teams qualify; raise to 75 for higher standards
Good	55	55-69	Adjust in tandem with Excellent threshold
Average	45	45-54	This is the "baseline" band; keep centered around 50
Below Average	30	30-44	Raise to 35 for earlier intervention; lower to 25 for fewer alerts
Needs Attention	0	0-29	Scores below this threshold require immediate focus

Part 2: Weight Preset Configurations

1. Balanced (Default)

Weights: CSS 50% | TRS 35% | PGS 15%

Best for: Most organizations with no specific evaluation bias

What it emphasizes: Equal consideration of where teams are today AND how they're improving

When to use:

- Standard operational assessments
- Organizations new to the CHS system
- When you want a fair balance between rewarding high performers and improving teams

Formula: CHS = 0.50 x CSS + 0.35 x TRS + 0.15 x PGS

2. Snapshot Focus

Weights: CSS 65% | TRS 25% | PGS 10%

Best for: Point-in-time audits, compliance assessments, executive reporting

What it emphasizes: Where teams stand RIGHT NOW, with less credit for trajectory

When to use:

- Annual compliance reviews
- Due diligence assessments
- When current capability matters more than improvement direction
- Reporting to stakeholders who need "current state" answers

Formula: CHS = 0.65 x CSS + 0.25 x TRS + 0.10 x PGS

3. Growth Focus

Weights: CSS 40% | TRS 45% | PGS 15%

Best for: Transformation programs, improvement initiatives, change management

What it emphasizes: Teams that are improving fastest, even if they started lower

When to use:

- During organizational transformations
- When rewarding progress is a strategic priority
- Agile maturity improvement programs
- Situations where starting position shouldn't penalize teams

Formula: CHS = 0.40 x CSS + 0.45 x TRS + 0.15 x PGS

---

4. Peer Comparison

Weights: CSS 45% | TRS 30% | PGS 25%

Best for: Competitive environments, cross-team benchmarking, relative performance evaluation

What it emphasizes: How teams compare to similar peers who started at the same level

When to use:

- When relative performance is the key metric
- Identifying which teams are "beating expectations" given their starting point
- Reducing the penalty for teams that inherited challenging situations

Formula: CHS = 0.45 x CSS + 0.30 x TRS + 0.25 x PGS

Part 3: Decision Framework for Threshold Adjustment

Use this flowchart to determine whether and how to adjust your thresholds:

```
START: Run assessment with default thresholds (70/55/45/30)
        |
        v
Are category distributions appropriate for your organization?
        |
        +-- YES --> Keep defaults. Review quarterly.
        |
        +-- NO --> What's the problem?
                   |
                   +-- Too many teams in "Excellent" (>25%)
                   |   |
                   |   v
                   |   Consider RAISING excellent threshold to 75
                   |   This preserves meaning of "excellent"
                   |
                   +-- Too many teams in "Needs Attention" (>20%)
                   |   |
                   |   v
                   |   First check: Is this accurate? If teams ARE strugg
                   |   If threshold is too sensitive: LOWER below-average
                   |
                   +-- Almost no teams in "Excellent" (<5%)
                   |   |
                   |   v
                   |   Consider LOWERING excellent threshold to 65
                   |   But first check: maybe your org needs to improve?
                   |
                   +-- Most teams clustered in "Average" (>50%)
                   |   |
                   |   v
                   |   This might be correct! Average of 50 is the baseli
                   |   Consider using Growth Focus preset to differentiat
                   |
                   +-- Categories don't match business expectations
                       |
                       v
                       Consider:
                       - Your organization's maturity level
                       - Industry benchmarks
                       - Historical score distributions
                       - Whether expectations are realistic
```

Key Principle

Adjust thresholds based on **multiple assessment periods**, never a single snapshot. Score distributions should be evaluated over at least 3 assessment cycles before considering threshold changes.

Part 4: Risk/Benefit Tradeoffs

CSS Weight (Current State Score)

If You...	Benefit	Risk
Increase (>0.50)	Rewards teams with strong current practices	Penalizes improving teams; may discourage transformation efforts
Decrease (<0.50)	Rewards improvement over position	May undervalue teams maintaining excellence; ceiling effect for top performers

Signs you should increase: Stakeholders primarily care about current capability; compliance requirements focus on "now"

Signs you should decrease: Organization is in transformation mode; you want to encourage struggling teams

TRS Weight (Trajectory Score)

If You...	Benefit	Risk
Increase (>0.35)	Strongly rewards improvement	May overweight short-term changes; volatile scores period-to-period
Decrease (<0.35)	More stable scores over time	Reduces incentive to improve; may feel unfair to improving teams

Signs you should increase: Improvement culture is a priority; teams have clear improvement goals

Signs you should decrease: Short assessment periods; high natural variability in your data

PGS Weight (Peer Growth Score)

If You...	Benefit	Risk
Increase (>0.15)	Fairer comparison accounting for starting point	Requires sufficient teams (20+); adds complexity to interpretation
Decrease (<0.15)	Simpler, more direct interpretation	May penalize teams that inherited challenging situations

Signs you should increase: Large variation in team starting points; want to reward "beating expectations"

Signs you should decrease: Small organization (<20 teams); prefer straightforward current + trajectory scoring

---

Excellent Threshold (default: 70)

If You...	Benefit	Risk
Raise to 75	Preserves exclusivity of "Excellent" rating	May demotivate high performers who miss the threshold
Lower to 65	More teams achieve recognition	Dilutes meaning of "Excellent"; may reduce improvement motivation

---

Below Average Threshold (default: 30)

If You...	Benefit	Risk
Raise to 35	Earlier intervention for struggling teams	More false positives; may create alarm fatigue
Lower to 25	Fewer teams flagged; focus on truly critical cases	May miss teams that need help; later intervention

Part 5: Minimum Prerequisites for Meaningful Scores

CHS scores require minimum data thresholds to be reliable. Understand what happens when these aren't met:

Requirement	Minimum	Recommended	What Happens If Not Met
Teams for PGS	20 teams	50+ teams	PGS unavailable; system uses 2-component model (CSS + TRS only)
Teams per baseline group	5 teams	10+ teams	Groups are merged with adjacent groups; may reduce PGS precision
Measurement periods	4 weeks	12+ weeks	TRS unreliable; short-term noise dominates trajectory calculation
Indicator coverage	70%	90%+	Available indicators are reweighted; results flagged as "partial coverage"
Historical data	8 weeks	24+ weeks	TRS weight automatically reduced; "provisional" flag applied
Data freshness	2 weeks	Weekly	Stale data warning; scores may not reflect current state

What "2-Component Model" Means

When you have fewer than 20 teams, PGS cannot be calculated. The system automatically redistributes the PGS weight proportionally:

- **With 3-component:** CHS = 0.50 x CSS + 0.35 x TRS + 0.15 x PGS
- **With 2-component:** CHS = 0.59 x CSS + 0.41 x TRS

This maintains the relative importance of CSS vs TRS while excluding peer comparison.

Part 6: Common Adjustment Scenarios

Scenario 1: New Deployment with Small Organization (15 teams)

Problem: PGS is unavailable because you have fewer than 20 teams

What the system does: Automatically uses 2-component model (CSS: ~59%, TRS: ~41%)

Recommendation:

- Accept the 2-component model for now
- Consider explicitly setting weights to CSS: 60%, TRS: 40%
- Plan to enable full CHS when you reach 20+ teams

What NOT to do: Don't try to force PGS calculation with too few teams - the peer comparisons would be statistically meaningless.

Scenario 2: Post-Acquisition Integration

Problem: You've merged teams from two organizations with different baseline populations

Symptoms:

- Acquired teams score very differently from existing teams
- Historical comparisons seem invalid
- Score distributions are bimodal (two peaks)

Recommendation:

1. Mark the first assessment post-acquisition as a "new baseline"
2. Consider running separate assessments for each population initially
3. Recalibrate baseline norms using the combined population
4. Use "Snapshot Focus" preset for the first 2-3 assessment periods
5. Gradually transition to "Balanced" once populations stabilize

Warning: Historical TRS comparisons may be invalid for 6-12 months post-acquisition.

Scenario 3: Seasonal Business Patterns

Problem: Year-end rush, holiday periods, or product launches affect Q4 scores differently than Q1-Q3

Symptoms:

- Predictable score dips in certain quarters
- Teams show "declining" trajectory during known busy periods
- Assessment timing affects results significantly

Recommendation:

1. Consider period-specific baseline norms (if data volume supports it)
2. Exclude known anomalous periods from TRS calculation
3. Increase TRS weight during recovery periods to capture rebound
4. Document seasonal patterns in assessment reports

Alternative: Compare scores to the same period in previous years rather than sequential periods.

Scenario 4: High-Maturity Organization

Problem: Most teams score well (60-75 range), creating score compression at the top

Symptoms:

- Few meaningful differences between top-performing teams
- "Excellent" category feels arbitrary
- High performers feel they "can't improve"

Recommendation:

1. Raise Excellent threshold to 75 or even 80
2. Consider using PGS more heavily (Peer Comparison preset)
3. Implement ceiling effect messaging for teams scoring 80+
4. Focus dashboard on "maintaining excellence" rather than improvement for top teams

Communication: Help high-performing teams understand that a CHS of 48-55 when starting at 80+ represents "sustained excellence" - holding steady at a high level is itself an achievement.

Scenario 5: Improvement Culture Initiative

Problem: Organization wants to emphasize and reward improvement, but current scoring feels static

Symptoms:

- Leaders ask "why don't improving teams get more recognition?"
- Teams focus on maintaining position rather than improving
- Feedback suggests scoring doesn't reflect effort

Recommendation:

1. Switch to Growth Focus preset (CSS: 40%, TRS: 45%, PGS: 15%)

2. Communicate the change clearly: "We're emphasizing trajectory"

3. Set expectations that scores will shift for some teams

4. Run both presets in parallel for one period to show impact

Alternative: Keep Balanced preset but create a separate "Improvement Leaders" recognition based on TRS alone.

Scenario 6: Teams Gaming the System

Problem: Some teams appear to be optimizing for metrics rather than real improvement

Symptoms:

- Rapid score jumps without visible practice changes
- Scores drop when data collection methods tighten
- Teams focus on easily-measured indicators

Recommendation:

1. Review indicator definitions for loopholes

2. Increase the number of required indicators

3. Apply stricter indicator coverage requirements (raise from 70% to 85%)

4. Consider adding qualitative review for large score jumps

Warning: Gaming often indicates that incentives are tied too tightly to scores. Consider whether scores are being used appropriately.

Part 7: Warning Signs and Guardrails

Things You Should NEVER Do

1. **Never set CSS weight below 30%**

   - Why: Current state must always matter. Ignoring where teams ARE today produces misleading scores.
   - Risk: A team could score "Excellent" while having poor current practices.

2. **Never set TRS weight below 15%**

   - Why: Trajectory adds essential information that pure snapshots miss.
   - Risk: Teams have no incentive to improve; scores become static.

3. **Never set PGS weight above 30%**

   - Why: PGS is inherently noisier than CSS and TRS.
   - Risk: Scores become volatile and harder to interpret.

4. **Never adjust thresholds based on a single assessment period**

   - Why: Single periods have natural variation; changes may be noise.
   - Wait: Minimum 3 assessment periods before considering threshold changes.

5. **Never customize parameters without documented rationale**

   - Why: Future administrators need to understand why settings were changed.
   - Do: Create a dated change log with reasoning for any adjustment.

6. **Never hide poor scores by raising thresholds**

   - Why: This masks real problems and delays necessary intervention.
   - Instead: Address the root causes; consider if expectations are realistic.

Healthy Parameter Ranges

Parameter	Minimum Safe Value	Maximum Safe Value
CSS weight	0.30	0.70
TRS weight	0.15	0.50
PGS weight	0.00	0.30
Excellent threshold	65	80
Good threshold	50	65
Average threshold	40	50
Below Average threshold	25	40

When to Seek Expert Help

Contact your statistical support team or methodology experts if:

- You want to change multiple parameters simultaneously

- Score distributions don't make sense after 3+ assessment periods

- You're considering weights outside the safe ranges

- You need to justify parameter choices to auditors or regulators

- Sensitivity analysis shows high volatility across configurations

---

Quick Reference Card

Default Configuration (Recommended for Most Organizations)

```
Weights:
  CSS (Current State):     50%
  TRS (Trajectory):        35%
  PGS (Peer Growth):       15%

Thresholds:
  Excellent:    >= 70
  Good:         >= 55
  Average:      >= 45
  Below Avg:    >= 30
  Needs Attn:   < 30


Formula: CHS = 0.50 x CSS + 0.35 x TRS + 0.15 x PGS
```

When to Consider Each Preset

If your priority is...	Use this preset
Balanced evaluation	Balanced (default)
Current capabilities matter most	Snapshot Focus
Rewarding improvement	Growth Focus
Fair comparison by starting point	Peer Comparison

Minimum Data Requirements

- Teams for full CHS: 20+

- Teams per peer group: 5+

- Weeks of data for TRS: 8+

- Indicator coverage: 70%+