Asia Pacific Science Press

Asia Pacific Economic and Management Review
ISSN(O):3005-9275
ISSN(P):3005-9267
Vol. 3 No.5 (2026)

Validating a DEA-to-fsQCA Framework for the Configurational Antecedents of R&D Team Innovation Efficiency Using a Known Data-Generating Process

Hongyu Cao[1], Weixiang Gan[1,2]*

1. Langfang Zhenguigu Technology Co., Ltd., Beifeng Road, Guangyang District, Langfang, Hebei, 065000, China
2. Lincoln University College, Petaling Jaya, Selangor Darul Ehsan 47301, Malaysia*

Corresponding author: Weixiang Gan, pdf.weixiang@lincoln.edu.my; Gabriel1995@qq.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: This study evaluates a two-stage analytical framework using controlled simulations with a known data-generating process. Studies of efficiency antecedents commonly use variable-oriented models to estimate net effects. These estimates do not directly represent conjunction, equifinality, or asymmetry. The framework separates efficiency measurement from configurational attribution. An input-oriented Banker-Charnes-Cooper data envelopment analysis (DEA-BCC) model provides efficiency scores for outcome calibration, and fuzzy-set qualitative comparative analysis (fsQCA) identifies combinations of psychological and situational conditions associated with the calibrated outcome. The simulation uses random seed 20260907 to generate 642 members across 160 R&D teams nested in 25 firms. The configurations underlying efficiency are therefore known. Mean raw BCC efficiency is 0.551, and 13 teams lie on the frontier. The strongest single condition accounts for 16.9% of the variation in efficiency, indicating limited explanatory power at the bivariate level. No condition or its negation meets the conventional threshold for necessity; the highest consistency is .757. Three sufficient configurations yield solution consistency of .887 and coverage of .571. Each of the 15 solution terms across six robustness variants contains knowledge sharing, work engagement, or both. Recovery of the generating structure is partial, with literal-level precision of .62 and recall of .86. One generating path is recovered exactly only after the efficiency model is changed. Knowledge sharing appears in all three configurations, including one whose generating counterpart does not contain it. The findings have two methodological implications. Simar and Wilson bias correction improves the point estimate but weakens the rank ordering used in calibration: the Spearman correlation with true efficiency falls from .521 to .337. Bounding the outcome at $\theta = 1$ concentrates cases at the frontier and depresses the proportional reduction in inconsistency (PRI) near the crossover. For applications to field data, the findings favor calibrating raw efficiency scores, reporting corrected intervals to describe estimation uncertainty, and selecting the PRI threshold with reference to the calibrated distribution.

Keywords: Data Envelopment Analysis; Fuzzy-Set Qualitative Comparative Analysis; Methodological Validation; Simulated Data with Known Ground Truth; Calibration; Innovation Efficiency

Published: Oct 22, 2026

DOI: https://doi.org/10.62177/apemr.v3i5.1771

1. Introduction

R&D teams with similar staffing, budgets, and observation periods can produce different numbers of invention patents and

completed new-product projects. Research in innovation management has addressed this variation by predicting output levels from R&D intensity, expenditure, and staffing, or by predicting innovative behavior from psychological capital and work engagement. Neither approach directly evaluates output relative to the inputs used to produce it.

Farrell (1957) defined productive efficiency as a unit's distance from the frontier attainable by comparable units. Charnes et al. (1978) and Banker et al. (1984) developed this approach into data envelopment analysis (DEA). DEA uses linear programming to construct a piecewise-linear frontier from the observed sample and evaluates each unit by its radial distance from that frontier. An input-oriented efficiency score of 0.551 indicates that a team could theoretically produce its observed patent and project outputs using 55% of its personnel and expenditure.

Distance from the frontier measures the scope for efficiency improvement but does not explain the conditions associated with a team's position. Different combinations of conditions may yield the same efficiency level, and the configurations associated with low efficiency need not mirror those associated with high efficiency. Conjunction, equifinality, and asymmetry concern relationships among combinations of conditions and are not directly represented by net regression coefficients, which describe average effects while holding other variables constant. In the simulated data, knowledge sharing is the strongest single condition and accounts for 16.9% of the variation in efficiency (Spearman $\rho = .411$). This limited bivariate association motivates a configurational second stage. Set-theoretic methods treat cases as combinations of memberships and represent solutions through intersection, union, and negation (Fiss, 2011; Greckhamer et al., 2018; Ragin, 2000, 2008; Schneider & Wagemann, 2012).

The ability-motivation-opportunity (AMO) framework links efficiency to the conditions that may contribute to it. The framework explains performance through the combined influence of what people can do, what they are willing to do, and what their work environment permits (Appelbaum et al., 2000; Blumberg & Pringle, 1982). Its multiplicative logic can be expressed through intersections of sets. In this study, psychological capital represents ability, work engagement represents motivation, and psychological safety, inclusive leadership, knowledge sharing, and an innovation-supportive climate represent opportunity.

Combining DEA and fsQCA raises methodological questions that field data alone cannot resolve. One concerns whether the outcome should be calibrated from raw DEA scores or from Simar and Wilson bias-corrected scores, which provide improved point estimates but alter the ranking of units. Another concerns how closely a recovered solution reflects the generating structure rather than the choice of calibration anchors, frequency threshold, or returns-to-scale assumption. These questions require knowledge of the configurations that generate efficiency, which field studies seek to identify. We therefore apply the full analytical framework to a controlled simulation with a specified data-generating process. The simulation comprises 160 teams nested in 25 firms, 642 members, and six conditions. Its efficiency engine is the maximum of three paths: A·M·S, L·K·C, and M·K·~L. All data are simulated, so the findings do not provide evidence about real R&D teams, firms, or populations. Reliability and agreement indices describe simulated item responses calibrated to ranges typical of published studies; they do not establish scale validity. The three paths are formulated as propositions P1 to P3 and inform directional expectations for counterfactual analysis. They are not hypotheses to be tested.

This study makes three methodological contributions. First, it specifies the decisions required in a two-stage DEA-to-fsQCA analysis: model orientation, returns to scale, percentile anchors, treatment of memberships equal to .50, frequency, consistency and PRI thresholds, and directional expectations. Second, it examines the choice of efficiency measure for calibration. Bias correction improves point estimation but weakens rank recovery. The Spearman correlation with true efficiency is .521 for raw scores and .337 for bias-corrected scores. These results favor calibrating raw scores and reporting corrected scores with confidence intervals to describe estimation uncertainty. Third, the study assesses recovery of the known generating structure. The main solution has consistency of .887 and coverage of .571. It embeds P1 in C1 = A·M·S·K, adding knowledge sharing to a generating path that does not contain it. The path L·K·C is recovered exactly under constant returns to scale, while M·K·~L appears only in robustness variants. The results indicate partial recovery and a systematic tendency to include knowledge sharing in every path.

2. Theoretical Background and Propositions

2.1 Innovation Efficiency as a Frontier Concept

Farrell (1957) distinguished technical efficiency from allocative efficiency. Technical efficiency evaluates performance relative to best practice among comparable units, whereas allocative efficiency also requires information on input prices. The technical component can therefore be applied to R&D teams without price data. Charnes et al. (1978) formalized this approach in the CCR model, which constructs a piecewise-linear cone around observed input-output vectors under constant returns to scale and measures efficiency as radial distance to the frontier. Banker et al. (1984) relaxed the returns-to-scale assumption in the BCC model. Its convex hull separates pure technical efficiency from inefficiency attributable to the scale of operation (Cooper et al., 2007).

Neither model requires a specified production function or a distributional assumption for the error term. Both accommodate multiple inputs and outputs through weights selected for each unit, avoiding an externally weighted composite index. Because the frontier is estimated from the same sample used to evaluate units, it is biased inward. Simar and Wilson (2000) developed a smoothed bootstrap to estimate this bias and construct confidence intervals.

Using estimated DEA scores in a conventional second-stage regression creates inferential problems. Simar and Wilson (2007) show that each score depends on a frontier constructed from the other units, producing complex dependence among estimation errors that diminishes only slowly as sample size increases. They also show that the concentration of scores at the upper boundary reflects finite-sample estimation bias rather than the data-generating process. A censored second-stage model is therefore misspecified, and neither Tobit nor OLS consistently estimates the relationship under these conditions. A set-theoretic second stage avoids this particular inferential problem but does not remove estimation error. Consistency and coverage describe subset relations in the calibrated data without relying on null hypotheses, sampling distributions, or p values. They therefore make no statistical inference that this dependence would invalidate. Each team's calibrated outcome membership still contains error from the estimated efficiency score. Set-theoretic measures do not report standard errors, so the residual dependence remains unquantified.

2.2 The AMO Framework

Blumberg and Pringle (1982) identified capacity, willingness, and opportunity as determinants of performance. Appelbaum et al. (2000) expressed these components as ability, motivation, and opportunity. AMO specifies how the components combine: its multiplicative formulation implies that a component close to zero limits performance regardless of the other two. Siemsen et al. (2008) tested this formulation and found stronger support for a constraining-factor model. In that model, knowledge sharing depends on the lowest level of motivation, opportunity, or ability, so improving a component that is not limiting has little effect. This minimum rule corresponds to the fuzzy-set intersection operator and provides a basis for applying AMO in the configurational stage.

The six conditions operationalize the AMO components at the team level. Psychological capital represents ability (Luthans et al., 2007; Avey et al., 2011). Work engagement represents motivation, following Kahn's (1990) conceptualization and Schaufeli et al.'s (2006) measurement approach. Four conditions represent opportunity: psychological safety (Edmondson, 1999), inclusive leadership (Carmeli et al., 2010), knowledge sharing (Van den Hooff & de Ridder, 2004), and innovation-supportive climate (Scott & Bruce, 1994). Including four opportunity conditions and one condition each for ability and motivation allows the analysis to examine potential substitution among distinct forms of opportunity.

2.3 The Configurational Perspective

Ragin's (2000, 2008) set-theoretic approach addresses conjunction, equifinality, and causal asymmetry, which a regression specification cannot directly represent together. Conjunction means that conditions operate in combination and that a condition's contribution depends on the configuration in which it appears. Equifinality allows different configurations to be sufficient for the same outcome. Causal asymmetry means that configurations sufficient for an outcome need not mirror those sufficient for its negation. Schneider and Wagemann (2012) formalize the analytical procedure, Greckhamer et al. (2018) specify reporting practices, and Du and Jia (2017) introduce the approach to Chinese-language management research.

Fiss (2011) distinguishes core conditions from peripheral conditions by comparing parsimonious and intermediate solutions. Core conditions appear in both solutions, whereas peripheral conditions appear only in the intermediate solution. Misangyi

and Acharya (2014) use this approach to examine whether conditions act as substitutes or complements. Substitutes can be sufficient in one another's absence, whereas complements must occur together. This distinction is relevant to the opportunity component of AMO. Although AMO's multiplicative formulation implies conjunction, it has mainly been estimated using interaction terms. Such specifications impose symmetry between presence and absence and restrict the analysis to the pathways represented by the chosen interaction structure.

2.4 Propositions as Directional Expectations

P1 combines psychological capital, work engagement, and psychological safety in a psychological-core path (A·M·S). Psychological resources and engagement sustain effort, while safety allows team members to pursue ideas that expose them to interpersonal risk. P2 combines inclusive leadership, knowledge sharing, and innovation-supportive climate in a contextual-enablement path (L·K·C). This path represents a complete opportunity structure. P3 combines engagement and knowledge sharing with the absence of inclusive leadership in a peer-driven path (M·K·~L). Following the substitution logic of Misangyi and Acharya (2014), peers provide coordination that would otherwise be supplied by a leader. The absence of inclusive leadership therefore helps define this path.

P1 through P3 serve as directional expectations rather than testable hypotheses. An fsQCA solution describes subset relations in calibrated data and does not provide a sampling distribution for rejecting an expectation. Directional expectations guide counterfactual simplification of logical remainders during minimization. The resulting solution terms are compared with the propositions without treating them as confirmed or rejected hypotheses.

The simulation engine also incorporates these three paths: E_set = max(A·M·S, L·K·C, M·K·~L). Because the generating configurations are known, recovery can be assessed by comparing the solution terms directly with those paths. Similarity to a proposition alone cannot establish that a solution term is correct, since the propositions are incorporated into the simulation design.

3. The Simulation Platform

3.1 Rationale for a Controlled Simulation

Field data alone cannot establish whether the configurations recovered by an analytical framework generated the observed outcome. Agreement between a solution and theory may reflect either successful recovery or the analyst's expectations. A controlled simulation provides benchmarks for each analytical stage. It allows us to assess whether DEA recovers the true ordering of teams, whether calibration preserves that ordering, and whether minimization recovers the generating paths. The parameters, scripts, and random seed 20260907 are fully specified, allowing readers to reproduce the member-level responses, calibrated memberships, and truth table. The analysis uses rectangular tables of inputs, outputs, and construct scores. Measured data can therefore be substituted without changing the DEA specification, calibration procedure, minimization procedure, or robustness variants.

The results do not provide evidence about real R&D teams. The coefficients do not describe a population, and the configurations do not establish practices that support managerial recommendations. The parameters were specified in advance to produce distributions within ranges reported in published team-level studies. They provide plausible conditions for evaluating the method rather than estimates from empirical observations.

3.2 Data-Generating Process

The platform has three layers.

The measurement layer generates 160 R&D teams nested in 25 firms, with three to five members per team and 642 members in total. It represents six constructs through 47 seven-point Likert items: psychological capital (A), work engagement (M), psychological safety (S), inclusive leadership (L), knowledge sharing (K), and innovation-supportive climate (C). Factor loadings range from .70 to .82, and target intraclass correlations place ICC(1) between .304 and .422. A structured correlation matrix produces moderate covariation among the opportunity constructs. Team-level scores are calculated as member means within the generating process and therefore contain aggregation error. Scores for a three-member team with dispersed responses are less precise than those for a five-member team with convergent responses. This imprecision is carried into calibration.

The causal layer combines the six construct memberships through E_set = max(A·M·S, L·K·C, M·K·~L). The tilde denotes absence. P1 through P3 identify these paths as directional expectations for recovery rather than hypotheses for testing. Taking the maximum makes the paths substitutes and introduces equifinality into the generating process. Firm fixed effects, team age, and a noise term influence the engine output, so identical configurational profiles can produce different levels of true capability.

The output layer translates inputs into outputs through a Cobb-Douglas frontier. R&D personnel is measured in full-time equivalents (FTE). Its elasticity is 0.42 in the invention-patent equation and 0.36 in the new-product-project equation. The corresponding elasticities for R&D expenditure are 0.48 and 0.52. Each observed output is the product of the frontier value, the capability multiplier from the causal layer, and an output noise term. Additional budget is assigned to teams with weak configurational profiles to create cases with large inputs and poor conversion efficiency. Without this adjustment, input size and efficiency would be approximately orthogonal by construction. Since all three layers are specified before estimation, a term that reproduces P1, P2, or P3 indicates recovery. Other terms reflect aggregation, efficiency estimation, calibration, or minimization.

3.3 Properties of the Simulated Responses

Table 1 summarizes the psychometric properties of the simulated item responses. Cronbach's alpha ranges from .806 to .913, composite reliability from .873 to .926, average variance extracted from .512 to .633, and r_wg from .839 to .897 (James et al., 1984). ICC(1) ranges from .304 to .422, and ICC(2) ranges from .636 to .746 (Bliese, 2000). Harman's single-factor test attributes 18.1% of the variance to the first factor (Podsakoff et al., 2003).

Table 1. Psychometric properties of the simulated item responses

Panel A. Measurement quality and convergent validity.

Construct	Source scale	Items	α	CR	AVE
Psychological capital (A)	Luthans et al. (2007)	12	0.913	0.926	0.512
Work engagement (M)	Schaufeli et al. (2006)	9	0.893	0.913	0.540
Psychological safety (S)	Edmondson (1999)	7	0.847	0.884	0.522
Inclusive leadership (L)	Carmeli et al. (2010)	9	0.908	0.924	0.575
Knowledge sharing (K)	Van den Hooff & de Ridder (2004)	6	0.857	0.894	0.583
Innovation-supportive climate (C)	Scott & Bruce (1994)	4	0.806	0.873	0.633

Panel B. Within-group agreement and reliability diagnostics.

Construct	rwg	ICC(1)	ICC(2)
Psychological capital (A)	0.897	0.422	0.746
Work engagement (M)	0.874	0.372	0.704
Psychological safety (S)	0.875	0.304	0.636
Inclusive leadership (L)	0.859	0.357	0.690
Knowledge sharing (K)	0.875	0.419	0.743
Innovation-supportive climate (C)	0.839	0.342	0.676

Note. N = 160 teams, 642 members, and 47 items. The source-scale column identifies the published instrument whose dimensional structure is reproduced by each generated item set. Item counts refer to simulated items and may differ from the length of the published instrument. The psychological capital set is a twelve-item analogue of the 24-item PCQ. α and CR are calculated at the member level; AVE is derived from principal-component loadings. rwg uses the scale-mean formulation with a uniform expected variance of 4.0. Harman's single-factor test attributes 18.1% of the variance to the first factor. All items use a seven-point Likert scale.

These indices describe simulated responses calibrated to ranges typical of published studies and do not establish scale validity. An alpha of .913 reflects the settings of the data generator. Matching published ranges allows the analytical framework to be assessed under plausible measurement conditions. The ICC(2) values of .636 for psychological safety and .676 for innovation-supportive climate were retained despite falling below the conventional .70 threshold, since such values are common in teams of three to five members.

Harman's test serves only as a calibration check. Simulated responses do not arise from respondents or a shared measurement procedure, so the test does not assess common-method bias in this setting. The value of 18.1% indicates that the specified correlation structure did not reduce the six constructs to one dominant factor. The simulated item sets reproduce the dimensional structure of the scales cited in Table 1 (Carmeli et al., 2010; Edmondson, 1999; Luthans et al., 2007; Schaufeli et al., 2006; Van den Hooff & de Ridder, 2004), but do not reproduce their wording. Innovation-supportive climate is deliberately simplified to one four-item dimension. The support-for-innovation and resource-supply subscales of Scott and Bruce's (1994) two-dimensional instrument are not modeled separately. No copyrighted item text is reproduced.

4. Analytical Strategy

4.1 A Two-Stage Design for Measurement and Attribution

The two analytical stages address different questions about the same 160 simulated teams. Data envelopment analysis (DEA) combines two inputs and two outputs into a relative efficiency score grounded in production theory. Fuzzy-set qualitative comparative analysis (fsQCA) uses the calibrated score as an outcome set and identifies conjunctions of conditions that form subsets of that outcome. This sequence avoids regressing efficiency scores that are bounded, serially correlated by construction, and estimated with bias that varies across units (Simar & Wilson, 2007). Calibration uses their relative positions within the (0, 1] interval.

4.2 DEA Specification

Efficiency is estimated using the input-oriented variable-returns-to-scale model of Banker et al. (1984). For each team o, the model is specified as follows:

$$\min_{\theta\ \lambda} \theta \quad \text{s.t.} \quad \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta\, x_{io}\ (\forall i), \quad \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro}\ (\forall r), \quad \sum_{j=1}^{n} \lambda_j = 1, \quad \lambda_j \ge 0, \tag{1}$$

The symbol x_{ij} denotes the ith input of team j. The inputs are R&D personnel in full-time equivalents and R&D expenditure. The symbol y_{rj} denotes the rth output, measured as invention patents or new-product projects. The sample contains $n = 160$ decision-making units. The weight λ_j represents team j's contribution to the reference set for team o, and $\theta \in (0,1]$ is the proportion of team o's input vector required to maintain its observed outputs. The convexity constraint $\sum_j \lambda_j = 1$ imposes variable returns to scale. Removing this constraint yields the constant-returns-to-scale CCR model of Charnes et al. (1978). The ratio $\theta_{\text{CCR}}/\theta_{\text{BCC}}$ measures scale efficiency and identifies inefficiency attributable to the scale of operation. With two inputs, two outputs, and 160 units, the sample satisfies the conventional size criteria $n \ge 3(m + s) = 12$ and $n \ge m \times s = 4$ (Cooper et al., 2007).

R&D personnel and budgets are often fixed within a planning cycle, making increased output from committed resources the most feasible adjustment. This rationale supports an output-oriented model, whereas equation (1) specifies an input-oriented model. Input orientation is appropriate when output commitments are fixed by contract or firm-level targets. It assesses how much of the current resource allocation could be reduced while maintaining the observed patent and project outputs. The two orientations do not coincide under variable returns to scale, so replication with an output-oriented model provides a relevant alternative specification.

A frontier estimated from a finite sample lies inside the true frontier, making each unit appear closer to the frontier than it actually is. The smoothed bootstrap of Simar and Wilson (2000) resamples the efficiency scores using a reflection kernel with bandwidth $h = 1.06\,\hat{\sigma} n^{-1/5}$. It draws $B = 1000$ pseudo-samples and produces the bias-corrected estimate $\tilde{\theta}_o = 2\hat{\theta}_o - B^{-1} \sum_b \hat{\theta}^*_{ob}$ with a 95% percentile interval.

Calibration assigns membership according to a score's position relative to the anchors rather than its accuracy as a point estimate. Bias correction improves the point estimate but reduces agreement with the known efficiency ranking: the Spearman

rank correlation is .521 for the raw score and .337 for the bias-corrected score. The raw BCC score is therefore used for calibration, while the bias-corrected score and its interval describe estimation uncertainty. One frontier team receives a bias-corrected estimate of −0.327, outside the admissible efficiency domain (0, 1], because the replication code clips the interval bounds but leaves the point estimate unbounded. This case is examined further in Section 5.1.

4.3 Calibration

Conditions and the outcome are converted into set memberships using Ragin's (2008) direct method. Log-odds interpolation connects three anchors: full non-membership e, the crossover point c, and full membership i. Membership is defined as follows:

$$f(x) = [1 + \exp(-\Lambda(x - c)/(i - c))]^{-1} \ (x > c) \ f(x) = [1 + \exp(\Lambda(c - x)/(c - e))]^{-1} \ (x \leq c) \quad (2)$$

The log-odds parameter is $\Lambda = \ln(.95/.05) \approx 2.944$, which sets $f(i) = .95$ and $f(e) = .05$. The anchors correspond to the 95th, 50th, and 5th percentiles of each variable's sample distribution and are reported in Table 2. Memberships of exactly .50 are increased by .001 to ensure an unambiguous truth-table assignment for each case (Fiss, 2011).

Table 2. Calibration anchors for the six conditions and the outcome

Set	Full non-membership (5th percentile)	Crossover (50th percentile)	Full membership (95th percentile)
Psychological capital (A)	3.665	4.694	5.753
Work engagement (M)	3.138	4.485	5.444
Psychological safety (S)	3.891	4.857	5.716
Inclusive leadership (L)	3.361	4.648	5.631
Knowledge sharing (K)	3.373	4.451	5.751
Innovation-supportive climate (C)	3.498	4.700	5.750
High innovation efficiency (raw BCC θ)	0.273	0.500	1.000

Note. Conditions are calibrated from construct scores aggregated to the team level, and the outcome is calibrated from the raw BCC efficiency score. Calibration follows Ragin's (2008) direct method with log-odds interpolation. Memberships of exactly 0.5 are increased by 0.001 to ensure that every case is assigned to a truth-table row (Fiss, 2011).

This calibration scheme has two limitations. Percentile anchors are sample-dependent: they define high membership relative to these 160 teams and would change with a different sample. A 90/50/10 specification is therefore included as a robustness variant. The outcome anchors are 0.273 / 0.500 / 1.000. The 95th percentile of the raw BCC score coincides with the theoretical maximum $\theta = 1$. Calibrated outcome memberships are consequently concentrated at the frontier, and the compressed upper tail depresses proportional reduction in inconsistency (PRI) for rows near the crossover. The PRI threshold is therefore set at .50 (Greckhamer et al., 2018; Schneider & Wagemann, 2012), and .70 is examined as a robustness variant.

4.4 Set-Theoretic Analysis

Necessity is assessed for each of the six conditions and its negation using consistency, relevance of necessity (RoN), and coverage. A condition is classified as necessary only when its consistency reaches .90 or above and its RoN is sufficient to rule out a trivially large set.

The truth table contains $2^6 = 64$ logically possible configurations. A row is coded as producing the outcome when it meets all three thresholds: consistency ≥ .80, PRI ≥ .50, and a frequency of at least two cases. Five rows qualify and contain 33 of the 160 teams. Sixteen configurations contain no cases, and eight contain a single team. Thus, 24 of the 64 rows (37.5%) enter minimization as logical remainders. This proportion reflects the limited coverage of the six-condition configuration space by 160 cases.

Directional expectations specify the presence of all six conditions, consistent with AMO reasoning that treats each condition

as performance-enhancing. This specification restricts the recovery of P3, which proposes a path involving the absence of inclusive leadership. Simplifying assumptions based on presence make this path harder to recover. Minimization produces three solutions. The conservative solution excludes logical remainders, the parsimonious solution uses all remainders, and the intermediate solution uses only remainders consistent with the directional expectations. The intermediate solution is reported here. Following Fiss's (2011) definition, conditions appearing in both the intermediate and parsimonious solutions are classified as core, while those appearing only in the intermediate solution are classified as peripheral.

The negated outcome is analyzed separately using its own truth table, thresholds, and reversed directional expectations. Configurations sufficient for an outcome's absence cannot be obtained by negating those sufficient for its presence (Greckhamer et al., 2018).

4.5 Software and Reproducibility

Simulation, DEA estimation, calibration, and figure generation are performed in Python 3.12 using NumPy, SciPy, pandas, and Matplotlib. The optimization problem in equation (1) is solved using SciPy's HiGHS interface. Necessity analysis, truth-table construction, and minimization are performed in R using the QCA package (Duşa, 2019; Thiem & Duşa, 2013). The random seeds are 20260907 for data generation and 42 for the bootstrap, enabling exact reproduction of every figure and table.

5. Results

5.1 Efficiency Estimates

Mean technical efficiency under the input-oriented BCC specification is 0.551 (SD 0.211) across the 160 simulated teams. The 13 teams (8.1%) with $\theta = 1$ lie on the frontier. Bootstrapping following Simar and Wilson (B = 1,000) reduces the mean to 0.489 (SD 0.201) and yields a median 95% confidence interval width of 0.082. Figures 1 and 2 show the efficiency distributions and rank-ordered bootstrap intervals, respectively. The downward correction is approximately six efficiency points, and the median interval spans eight points. These values are typical for a sample of this size and often conclude first-stage reporting in two-stage applications. Three additional properties of these estimates warrant closer examination.

Figure 1. Distribution of raw and bias-corrected BCC efficiency scores.

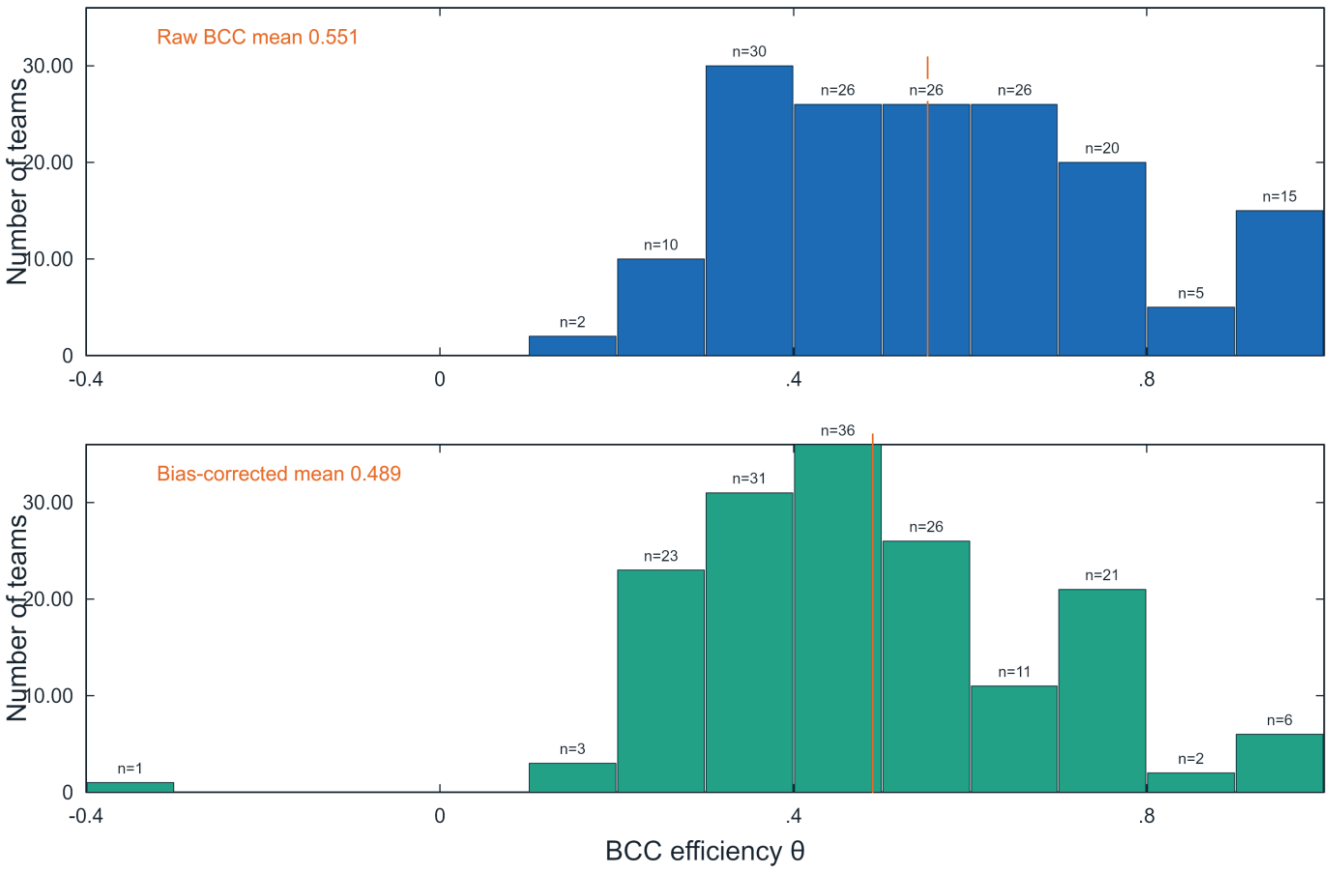
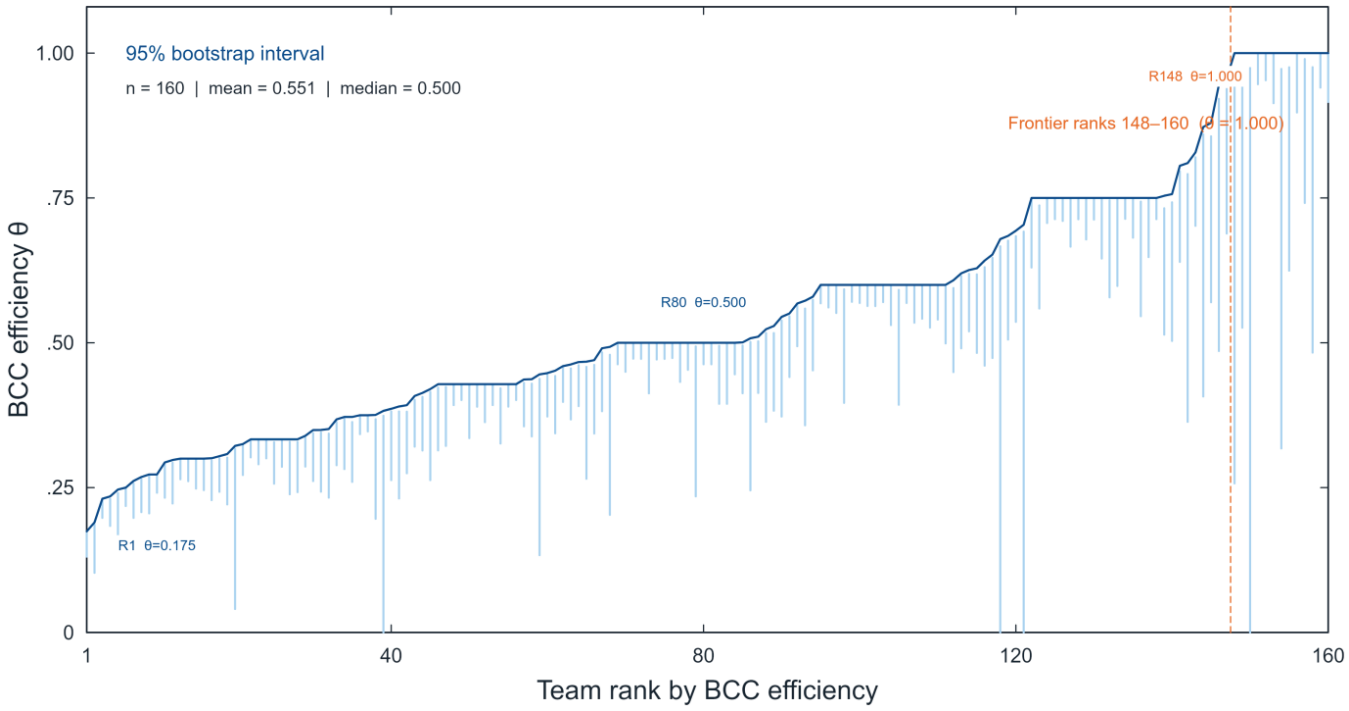


Figure 2. Rank-ordered raw BCC efficiency with 95% bootstrap intervals.



Estimation precision is lowest for teams on the frontier. The median interval width for the 13 frontier teams is 0.248, three times the value of 0.081 for the other 147 teams. The upper bounds for frontier teams range from 0.968 to 0.9997 and do not reach 1. Frontier membership is therefore the least certain classification in this sample. These same 13 teams define the full-membership anchor of the calibrated outcome, so the teams with the widest estimation intervals determine the scale used for all other teams in the second stage.

The bootstrap also produces an out-of-range estimate. Team 65 receives a bias-corrected score of -0.327, outside the admissible efficiency domain (0, 1]. This value arises from the estimator rather than a data error. The script clips the bootstrap interval bounds to the unit interval but leaves the bias-corrected point estimate unbounded, allowing the estimated bias to exceed the raw score. Subtracting a bootstrap bias estimate from the raw score (Simar & Wilson, 2000) does not guarantee a value within the efficiency domain. For a frontier team in this sample of 160 units, the adjustment produces a value below that domain. Reporting the estimate without clipping allows readers to assess this limitation.

Bias correction improves the point estimates but reduces the accuracy of the ranking. Rank correlation with the efficiency values specified in the data-generating process is .521 for the raw scores and .337 for the bias-corrected scores. The moderate correlation for the raw scores shows that the first stage recovers the true ordering imperfectly even before correction. Bias correction further reduces this agreement. Calibration depends on relative position because the direct method maps scores to memberships using percentile anchors. An order-preserving rescaling leaves the fuzzy sets unchanged, whereas a transformation that changes team rankings alters them. The raw score is therefore used in the second stage, while the bias-corrected score and its interval describe estimation uncertainty. Calibration using the corrected score is examined as robustness variant R5.

The constant-returns-to-scale assumption changes both efficiency levels and rankings. Mean CCR efficiency is 0.393, with a mean gap of 0.159 below the BCC scores. Mean scale efficiency θ_CCR/θ_BCC is .736 (SD .251). Figures 3 and 4 show the model comparison and the scale-efficiency distribution, respectively. Rank agreement between the models is weaker than linear agreement: Spearman ρ = .502, compared with a Pearson coefficient of .597. The two assumptions produce appreciably different rankings of the same 160 teams. This specification choice matters because the second stage depends on those rankings. The CCR score is therefore examined as robustness variant R6.

Figure 3. CCR efficiency against BCC efficiency scores.

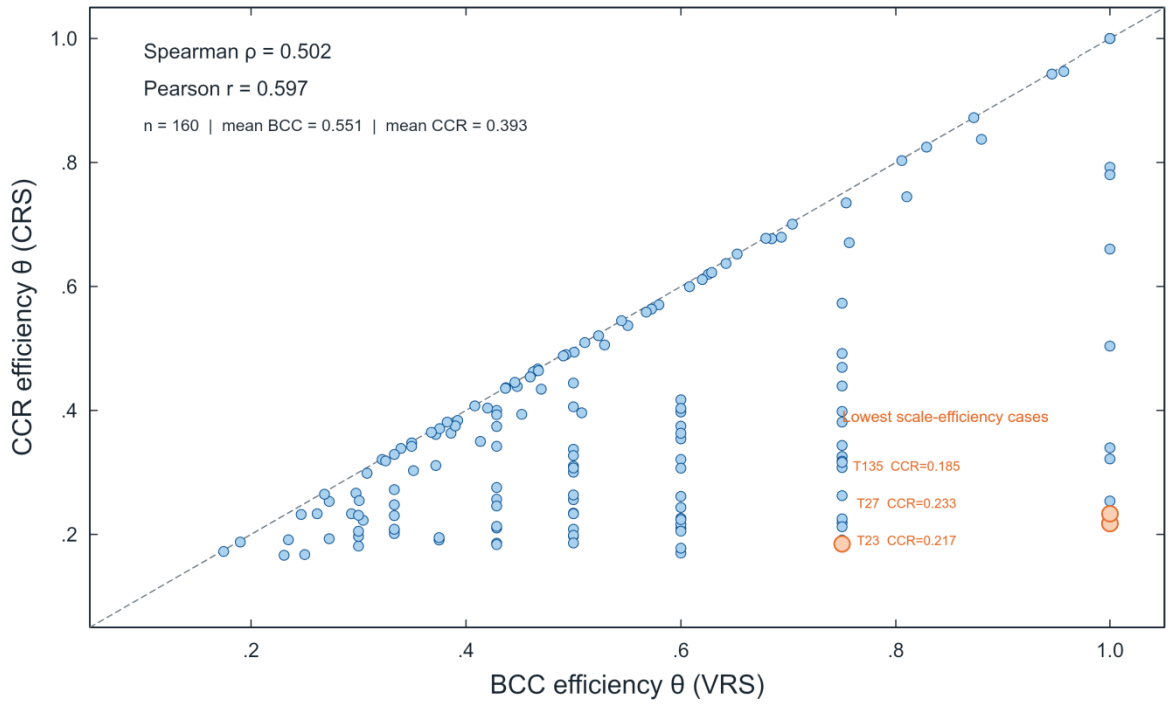
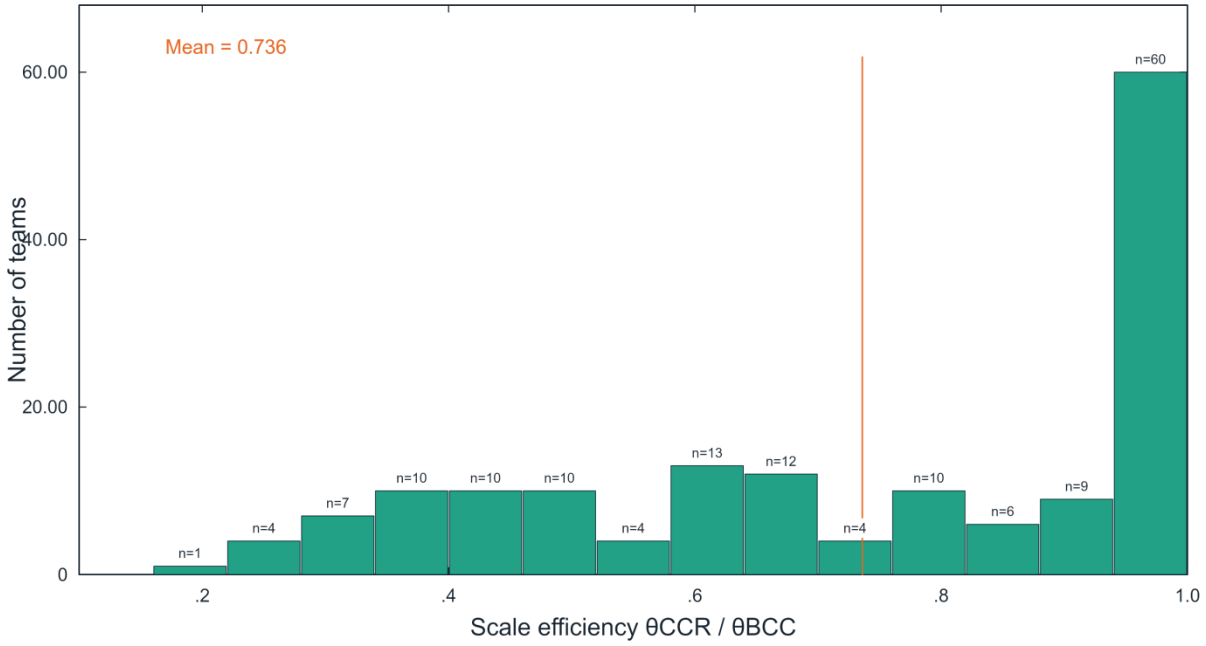


Figure 4. Distribution of scale efficiency (CCR/BCC).



Figures 5 and 6 show the bivariate relationships relevant to the second-stage analysis, and Table 3 reports the corresponding means, standard deviations, and correlations. Correlations among the six conditions range from .14 to .46 in Figure 5. These values indicate sufficient covariation for a configurational analysis while retaining distinctions among the six sets. Condition means range from 4.47 to 4.80 on the seven-point scale, with standard deviations of 0.59 to 0.72. No condition is concentrated at the scale ceiling to an extent that would compress its calibrated set. The square root of AVE ranges from .716 to .796 and exceeds every off-diagonal correlation in the matrix. Figure 6 ranks the conditions by their Spearman correlations with efficiency, consistent with the ranking preserved by calibration. Knowledge sharing has the highest correlation at .411 and accounts for 16.9% of the variation in efficiency. Inclusive leadership has the lowest correlation at .194. The remaining correlations are .332 for innovation-supportive climate, .308 for work engagement, .292 for psychological capital, and .210 for psychological safety. No single condition provides a strong account of the outcome, supporting the use of a

configurational second stage in place of a regression using the same six variables. The means and correlations in Table 3 describe simulated item responses calibrated to ranges commonly reported in published studies. They confirm that the simulation produced the intended correlation structure and do not establish the validity of the source scales.

Table 3. Descriptive statistics and correlations

Variable	M	SD	1	2	3	4	5	6
1. Psychological capital (A)	4.68	0.64	—					
2. Work engagement (M)	4.47	0.67	0.37***	—				
3. Psychological safety (S)	4.80	0.59	0.21**	0.14	—			
4. Inclusive leadership (L)	4.60	0.68	0.26***	0.32***	0.46***	—		
5. Knowledge sharing (K)	4.52	0.72	0.37***	0.41***	0.38***	0.43***	—	
6. Innovation-supportive climate (C)	4.68	0.72	0.22**	0.38***	0.26***	0.46***	0.41***	—
7. Innovation efficiency θ	0.55	0.21	0.30***	0.34***	0.21**	0.22**	0.42***	0.32***

Note. N = 160 teams. Entries below the diagonal are Pearson correlations. Variables 1 to 6 are construct scores aggregated to the team level on seven-point scales; variable 7 is the raw input-oriented BCC efficiency score. The square root of AVE ranges from 0.716 to 0.796 and exceeds every off-diagonal correlation, supporting discriminant validity. * p < .05, ** p < .01, *** p < .001.

Figure 5. Pearson correlations among the six conditions and BCC efficiency.

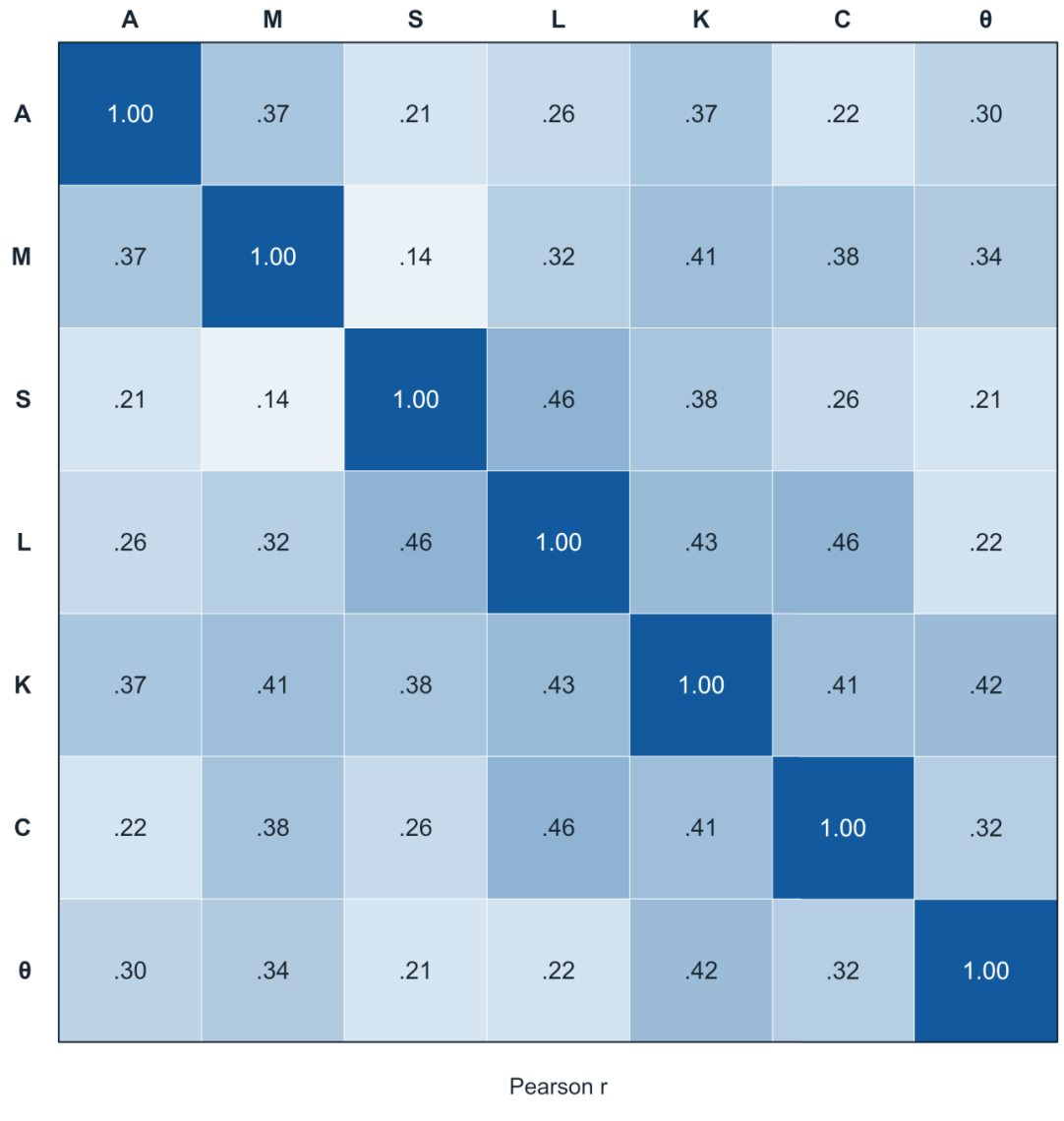
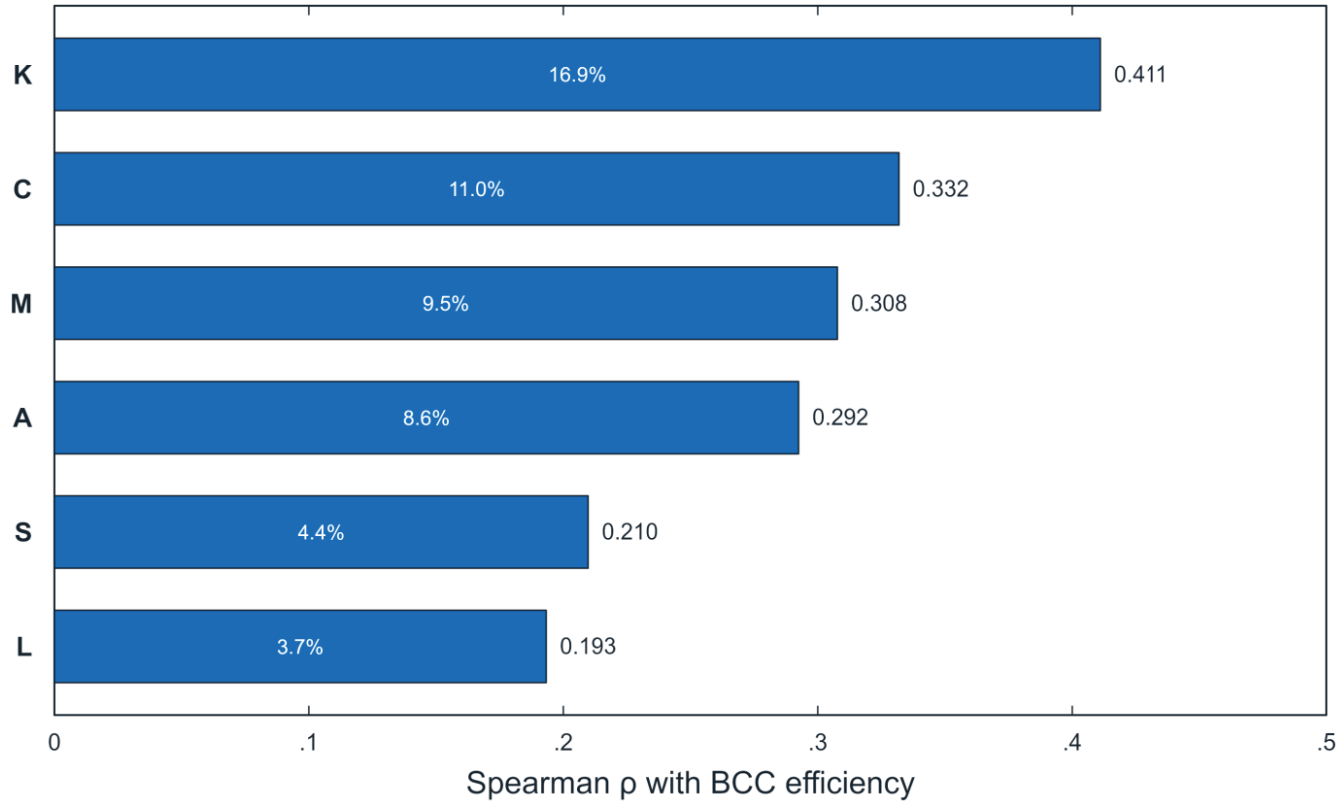


Figure 6. Spearman association of each condition with BCC efficiency.



5.2 Calibration

Figures 7 and 8 illustrate calibration. Figure 7 shows the direct-method curve (Ragin, 2008) for knowledge sharing. The anchors reported in Table 2 correspond to the 5th, 50th, and 95th percentiles of the raw team score and map to memberships of .05, .50, and .95. Log-odds interpolation with $\Lambda = 2.944$ determines membership between the anchors. The curve compresses differences in both tails and expands them around the crossover. Figure 8 shows the resulting distributions for all seven calibrated sets.

Figure 7. Direct-method calibration for knowledge sharing.

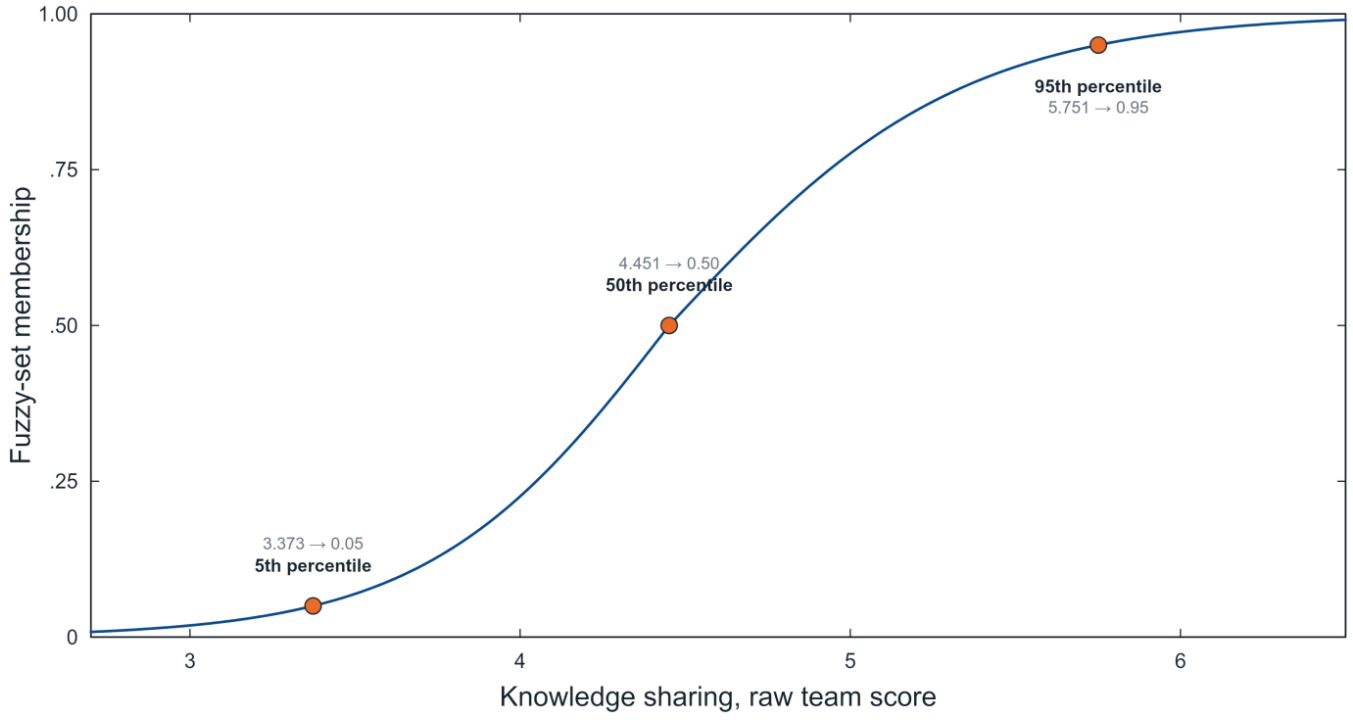
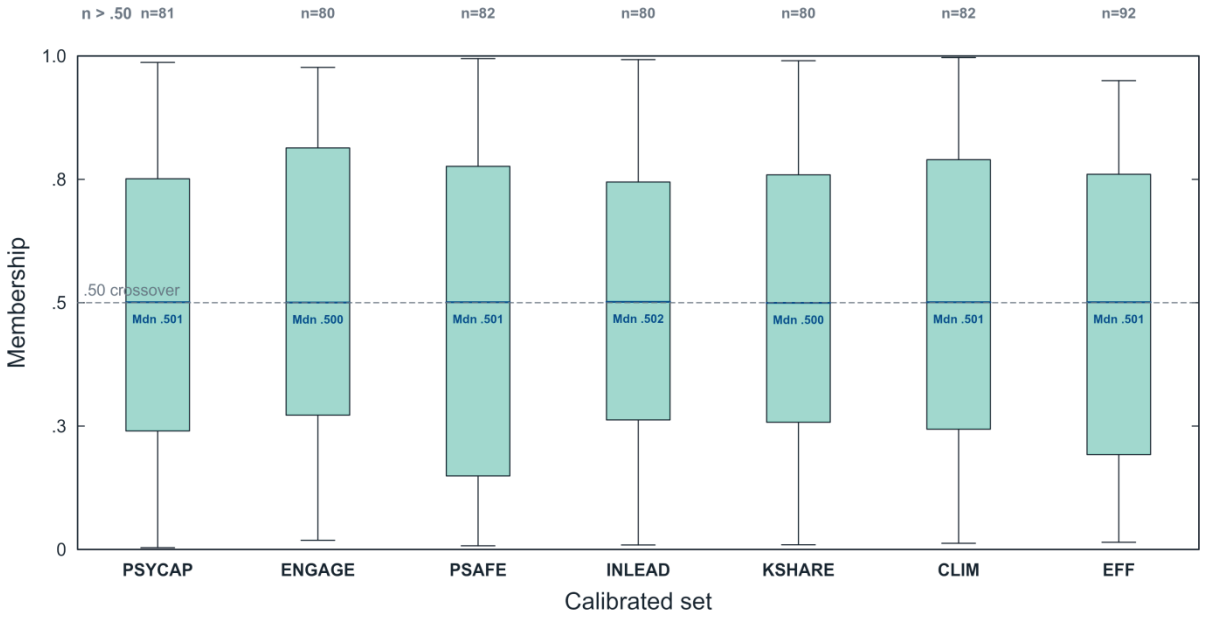


Figure 8. Distributions of calibrated memberships across conditions and the outcome.



Percentile anchors determine key features of these distributions. For each condition, approximately half of the 160 teams fall above the crossover because it is fixed at the median. This split reflects the calibration rule rather than a feature of the simulated responses. The outcome set differs at the upper end of its distribution. Its full-membership anchor is the 95th percentile of θ, which equals 1.000 because 13 teams have that score. All 13 frontier teams therefore receive a membership of exactly .950, and no team exceeds this value. The outcome distribution consequently has a point mass at its ceiling. Memberships of exactly .50 are increased to .501 (Fiss, 2011). The concentration at the ceiling depresses PRI for truth-table rows near the crossover, motivating the PRI threshold of .50.

5.3 Necessity

Figure 9 and Table 4 report necessity consistency, relevance of necessity, and coverage for the six conditions and their negations with respect to high efficiency and its negation. For condition presence, consistency ranges from .680 for psychological safety to .757 for knowledge sharing. Relevance of necessity ranges from .746 to .781, and necessity coverage ranges from .683 to .729. Consistency is lower for the negated conditions, ranging from .575 to .627. The pattern reverses for low efficiency: consistency ranges from .586 to .629 for condition presence and from .686 to .730 for condition absence.

Table 4. Analysis of necessary conditions

Panel A. Necessary conditions for high innovation efficiency.

Condition	Consistency	RoN	Coverage
Psychological capital (A)	0.719	0.780	0.711
Work engagement (M)	0.754	0.746	0.697
Psychological safety (S)	0.680	0.774	0.688
Inclusive leadership (L)	0.704	0.756	0.683
Knowledge sharing (K)	0.757	0.781	0.729
Innovation-supportive climate (C)	0.724	0.767	0.702
~Psychological capital (A)	0.613	0.705	0.592
~Work engagement (M)	0.575	0.736	0.596
~Psychological safety (S)	0.625	0.695	0.591
~Inclusive leadership (L)	0.627	0.727	0.618
~Knowledge sharing (K)	0.580	0.709	0.576
~Innovation-supportive climate (C)	0.586	0.707	0.578

Panel B. Necessary conditions for low innovation efficiency.

Condition	Consistency	RoN	Coverage
Psychological capital (A)	0.596	0.728	0.617
Work engagement (M)	0.628	0.694	0.607
Psychological safety (S)	0.586	0.738	0.620
Inclusive leadership (L)	0.629	0.731	0.638
Knowledge sharing (K)	0.592	0.706	0.596
Innovation-supportive climate (C)	0.591	0.710	0.599
~Psychological capital (A)	0.721	0.782	0.728
~Work engagement (M)	0.686	0.815	0.745
~Psychological safety (S)	0.705	0.755	0.697
~Inclusive leadership (L)	0.687	0.777	0.709
~Knowledge sharing (K)	0.730	0.811	0.759
~Innovation-supportive climate (C)	0.706	0.789	0.728

Note. N = 160. Panel A reports results for high innovation efficiency, and Panel B reports results for its negation. RoN denotes relevance of necessity. A condition is classified as necessary when consistency reaches .90. The highest observed value is 0.757, so no condition or negated condition meets this threshold. The tilde denotes set negation.

Figure 9. Necessity consistency for high and low innovation efficiency.

The highest consistency across the 24 tests is .757, below the conventional necessity threshold of .90 (Ragin, 2008; Schneider & Wagemann, 2012). No condition or negated condition is necessary for either outcome in these simulated data. Necessity tests assess individual conditions and do not capture multiplicative combinations. When the generating structure is multiplicative, a constraint on conversion may arise from several different conditions, so no single condition must be present in every high-efficiency case. Knowledge sharing and work engagement have the highest consistency values at .757 and .754, respectively. Knowledge sharing also has the highest necessity coverage at .729. The differences from innovation-supportive climate at .724 and psychological capital at .719 are too small to support a substantive ranking, and neither of the two highest values approaches the necessity threshold. Knowledge sharing and work engagement both recur in the sufficiency analysis, where their roles within configurations can be examined.

5.4 Sufficient Configurations for High Efficiency

Table 5 presents the truth-table rows that meet all three thresholds: raw consistency ≥ .80, PRI ≥ .50, and frequency n ≥ 2. Five rows qualify and together contain 33 of the 160 teams. Their consistencies range from .891 to .931, and their PRI values range from .568 to .755. The row with all six conditions present contains 18 of those 33 teams, with consistency .911 and PRI .755. The other four rows contain 6, 5, 2, and 2 teams. One of the two-case rows has the highest consistency at .931. Eight additional rows contain at least one case but fall below the frequency threshold, and 16 rows are logical remainders with no observed cases. Counterfactual analysis specified the presence of all six conditions as directional expectations.

Table 5. Truth-table rows meeting the consistency, PRI, and frequency thresholds

A	M	S	L	K	C	Cases	Consistency	PRI
1	1	1	1	1	1	18	0.911	0.755
0	1	1	1	1	1	6	0.896	0.600
1	1	0	1	1	1	5	0.891	0.568
1	1	1	0	1	0	2	0.914	0.594
1	1	1	0	1	1	2	0.931	0.671

Note. The table presents rows meeting the frequency threshold (n ≥ 2), consistency threshold (≥ .80), and PRI threshold (≥ .50). The 5 qualifying rows together contain 33 of the 160 teams. A membership threshold of .50 determines presence or absence: 1 = condition present, 0 = condition absent. A = psychological capital, M = work engagement, S = psychological safety, L = inclusive leadership, K = knowledge sharing, C = innovation-supportive climate.

Minimization of these rows yields the three terms presented in Table 6 and Figures 10 and 11. C1 (A·M·S·K) has consistency .910, PRI .775, raw coverage .464, and unique coverage .039. C2 (M·S·K·C) has consistency .901, PRI .759, raw coverage .480, and unique coverage .055. C3 (A·M·L·K·C) has consistency .896, PRI .732, raw coverage .443, and unique coverage .052. Solution consistency is .887, solution PRI is .743, and solution coverage is .571. PRI is .135 to .164 below raw consistency for the individual terms and .144 below raw consistency for the overall solution. This gap reflects outcome calibration. The full-membership anchor coincides with the theoretical maximum of 1, concentrating outcome memberships near the frontier and lowering PRI for cases close to the crossover.

Table 6. Configurations sufficient for high innovation efficiency

Condition	C1	C2	C3
Psychological capital (A)	●		●
Work engagement (M)	●	●	●
Psychological safety (S)	●	●	
Inclusive leadership (L)			●
Knowledge sharing (K)	●	●	●
Innovation-supportive climate (C)		●	●
Consistency	0.910	0.901	0.896
PRI	0.775	0.759	0.732
Raw coverage	0.464	0.480	0.443
Unique coverage	0.039	0.055	0.052

Note. N = 160. ● = core condition present, ⊗ = core condition absent, blank = condition may be present or absent. Overall solution consistency = 0.887; overall solution coverage = 0.571; overall PRI = 0.743. The frequency threshold is n ≥ 2, the consistency threshold is .80, and the PRI threshold is .50. Directional expectations for counterfactual analysis specified the presence of all six conditions. Every condition in a term is classified as core because the intermediate and parsimonious solutions coincide. Minimization yields 4 equivalent models. All four are listed in Section 5.4.

Figure 10. Configurations sufficient for high innovation efficiency.

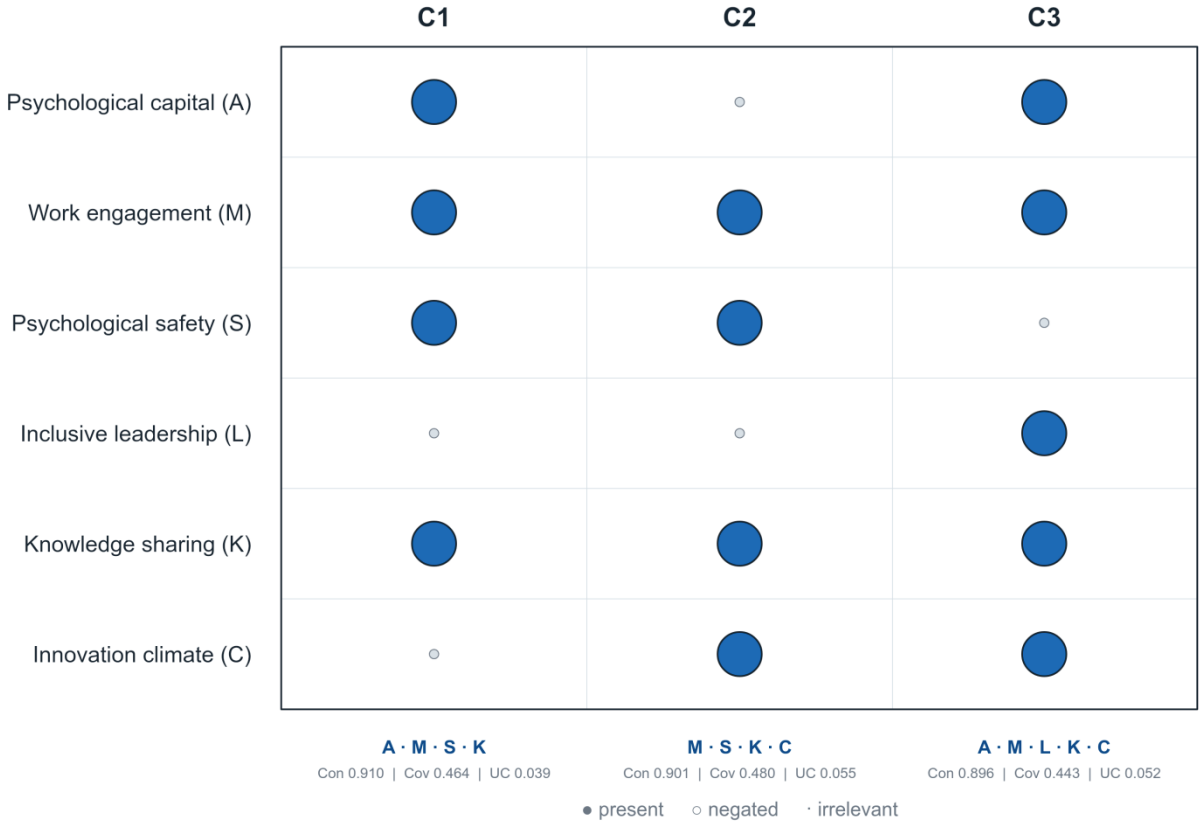
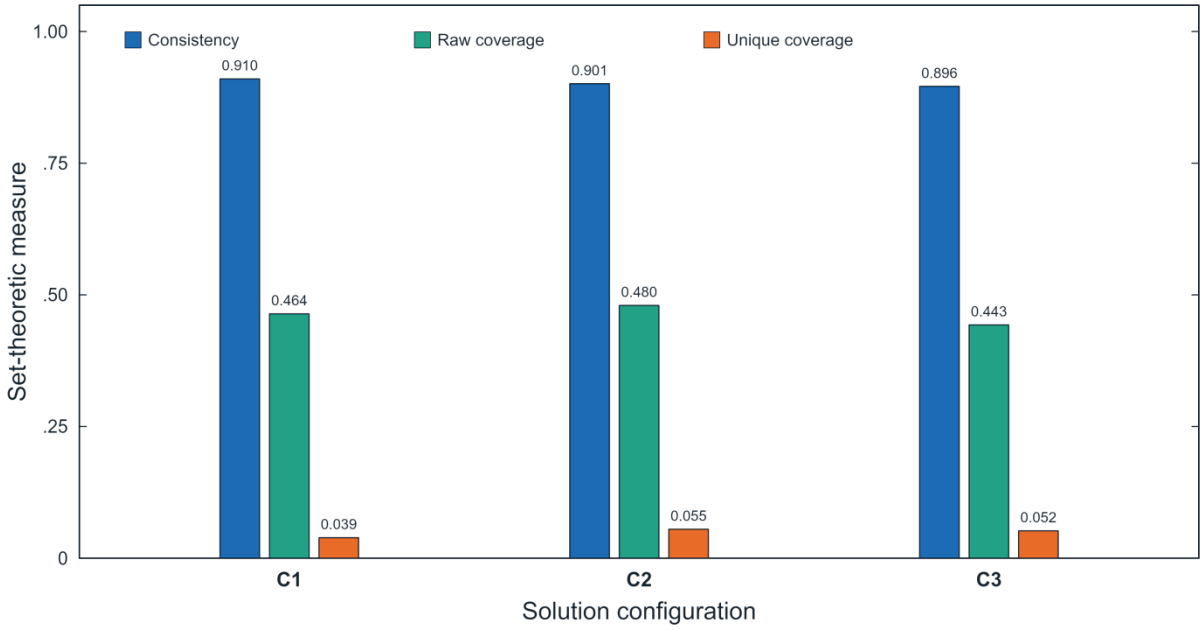


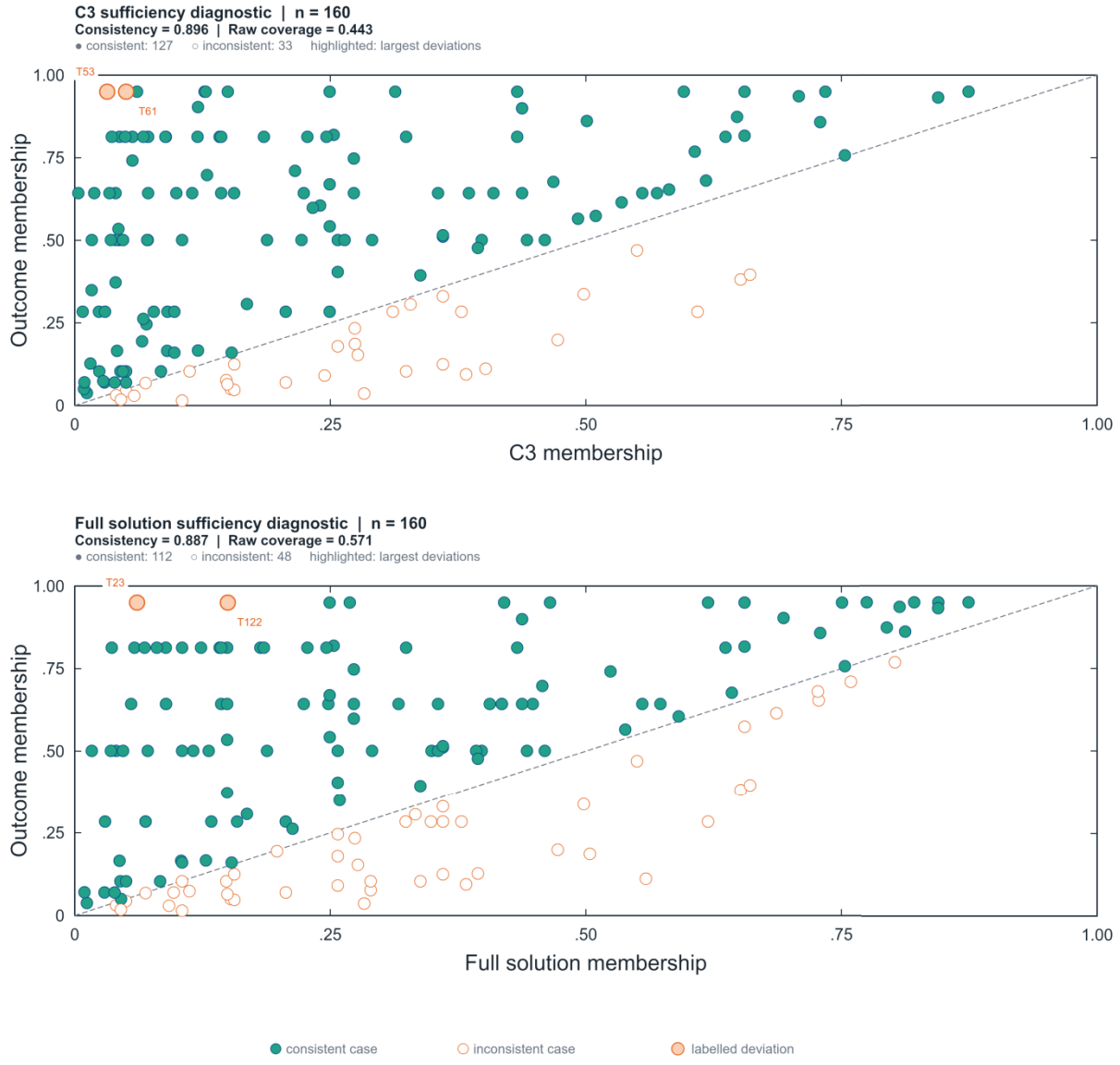
Figure 11. Term-level consistency, raw coverage, and unique coverage.



Figures 12 and 13 plot configuration membership against outcome membership for each term and the full solution, with a diagonal reference line. Cases above the diagonal have higher membership in the configuration than in the outcome and therefore count against sufficiency. These cases number 34 for C1, 36 for C2, 33 for C3, and 48 of 160 for the overall solution. Thus, term consistencies near .90 coexist with approximately one in five cases having lower outcome membership than configuration membership. The overall solution has consistency .887 and 48 such cases. Solution membership is the maximum membership across the three terms, so a case counts against sufficiency for the solution whenever it does so for at least one term. The 48 cases form the union of the sets of 34, 36, and 33 and are substantially fewer than their sum. Consistency summarizes these membership differences in a single value, whereas the plots show their magnitude and

distribution across the membership range.

Figure 12. Sufficiency diagnostics for configurations C1 and C2.



C1 combines psychological capital and work engagement with psychological safety and knowledge sharing. Its four literals include all three literals of P1 plus knowledge sharing, making it a literal-wise superset of that expectation rather than an exact match. C2 is the only term without psychological capital. Work engagement, psychological safety, knowledge sharing, and a supportive climate appear together in its absence. C3 contains the largest number of conditions. Its five literals include P2's L·K·C alongside psychological capital and work engagement.

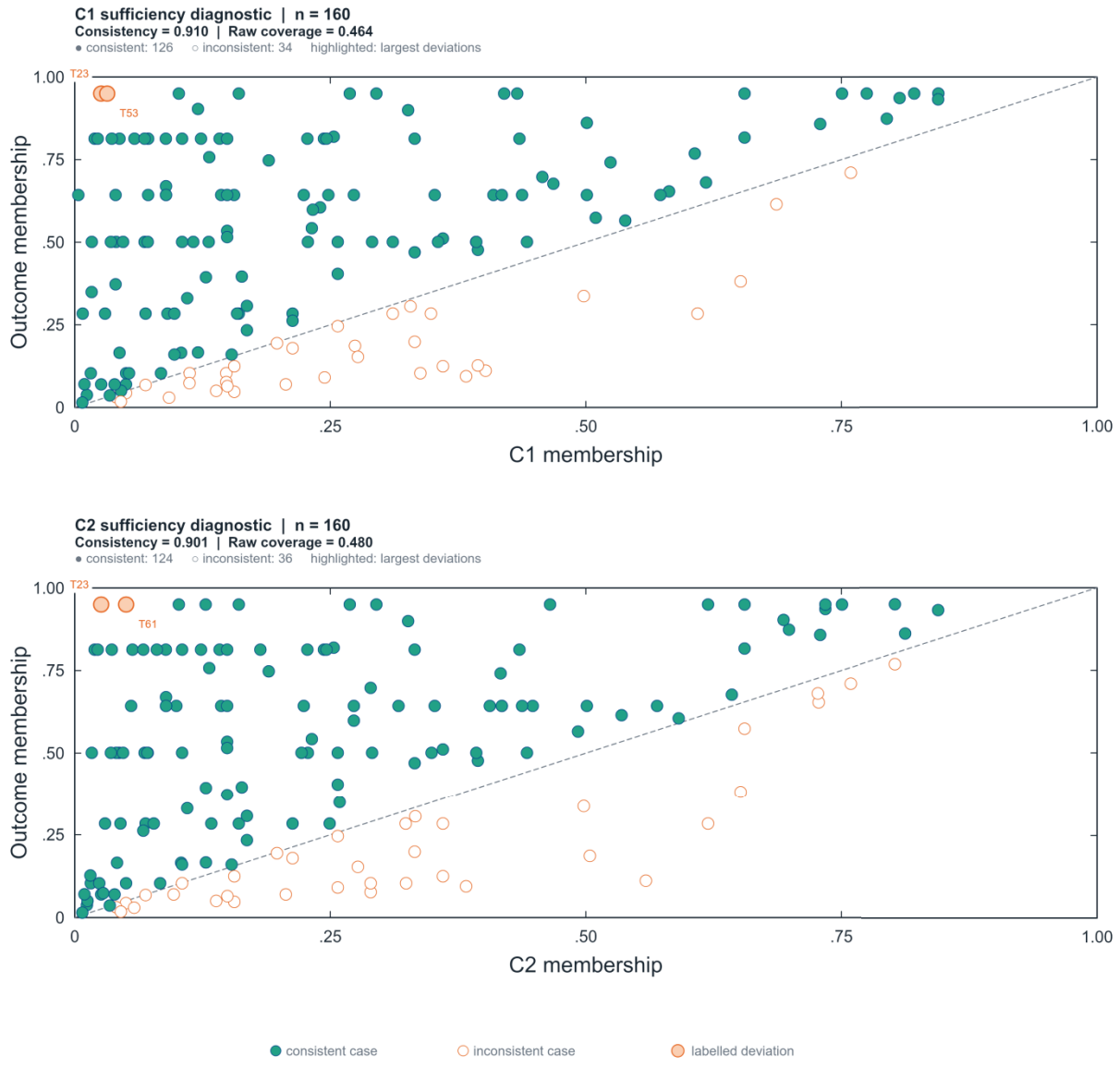
Unique coverages range from .039 to .055, compared with raw coverages of .443 to .480, and the three unique values sum to .146. The three paths therefore overlap substantially in the cases they account for. Equifinality here indicates that several routes cover many of the same teams. It does not divide the sample into three mutually exclusive groups, each covered by one route. Assigning each team to a single configuration would therefore misrepresent the coverage results.

Knowledge sharing and work engagement appear in all three terms. We describe a condition as ubiquitous within a solution when it appears as a literal in every term of that solution. This label describes recurrence within the solution and does not indicate causal importance. It differs from Fiss's (2011) core/peripheral distinction, which depends on whether a condition remains in the parsimonious solution rather than its frequency across terms.

The core/peripheral distinction provides no differentiation in this application. The intermediate and parsimonious solutions coincide for each model, so the intersection rule classifies every condition in every term as core. The uniform shading in

Table 6 reflects this coincidence and does not constitute a separate substantive finding.

Figure 13. Sufficiency diagnostics for configuration C3 and the full solution.



Minimization yields four equivalent intermediate models. All four share C2 and C3 and differ only in the first term: (1) A·M·S·K, (2) A·S·~L·K, (3) A·$_{L·K}$C, and (4) M·S·~L·K. Three of the four first terms contain ~L; model 1 does not. Knowledge sharing is the only literal common to all four, so ~L recurs in a majority of the alternatives without forming a shared structure. Model 1 is presented in the main text because its first term is the only one without a negated literal. This choice aligns the displayed solution with the directional expectations of presence used in the analysis. Selecting the model because it matches P1 would introduce circular reasoning because the generating process is known by design. All four models fit the data equally well.

Coverage defines the scope of the three terms. Solution coverage of .571 leaves approximately four-tenths of the membership in the high-efficiency outcome set outside their coverage, and the underlying truth-table rows contain 33 of 160 teams. The solution does not address the remaining teams.

5.5 The Negated Outcome

Repeating minimization for the negation of high efficiency with reversed directional expectations yields an overall solution consistency of .706. This is below the .75 minimum conventionally used to interpret a sufficiency solution (Schneider & Wagemann, 2012). The four terms in Table 7 are therefore exploratory and do not support any other conclusions in this paper. Overall coverage of .918 does not compensate for the low consistency. This combination indicates that the terms are

sufficiently broad to overlap most of the sample, including simulated teams with high membership in the efficiency set.

Table 7. Exploratory configurations sufficient for low innovation efficiency

Condition	D1	D2	D3	D4
Psychological capital (A)			⊗	
Work engagement (M)	⊗			
Psychological safety (S)			⊗	
Inclusive leadership (L)				●
Knowledge sharing (K)		⊗		
Innovation-supportive climate (C)				⊗
Consistency	0.745	0.759	0.785	0.844
PRI	0.556	0.594	0.600	0.645
Raw coverage	0.686	0.730	0.550	0.464
Unique coverage	0.058	0.083	0.019	0.023

Note. N = 160. The outcome is the negation of high innovation efficiency. ● = condition present, ⊗ = condition absent, blank = condition may be present or absent. Overall solution consistency = 0.706; overall coverage = 0.918. Consistency falls below the conventionally required range of .75 to .80, so these paths are reported as exploratory.

D1 represents the absence of work engagement and has consistency .745, PRI .556, raw coverage .686, and unique coverage .058. D2 represents the absence of knowledge sharing (.759, .594, .730, .083). D3 represents the joint absence of psychological capital and psychological safety (.785, .600, .550, .019). D4 combines inclusive leadership with the absence of an innovation-supportive climate (.844, .645, .464, .023). Three of the four term-level consistencies exceed .75, but the consistency of the combined solution falls below this threshold because the terms overlap substantially in the cases they cover.

These findings permit a structural interpretation without supporting a causal claim. D1 and D2 are single conditions, and D3 and D4 are two-condition expressions. By contrast, the high-efficiency terms in Table 6 are four- and five-condition conjunctions. The two solutions are not mirror images. This pattern represents causal asymmetry in a set-theoretic solution, which differs from reversing the sign of a coefficient (Fiss, 2011; Ragin, 2008). The asymmetry describes the relationship between the solutions; the present evidence does not establish how efficiency is lost. D4 has the highest term consistency of the four, but its PRI of .645 indicates that much of its membership is also consistent with the presence of high efficiency. It also belongs to a solution with consistency below the minimum threshold. No interpretation of leadership behavior is therefore based on this term.

5.6 Robustness

Table 8 and Figures 14 and 15 present six variants grouped by the analytical choice each modifies.

Table 8. Robustness checks across six analytic variants

Panel A. Analytic specification and outcome quality.

Variant	Specification	Rows	Terms	Consistency	Coverage
Main	Consistency .80, PRI .50, n ≥ 2, 95/50/5 anchors, raw BCC θ	5	3	0.887	0.571
R1	Consistency threshold raised to .85, PRI to .60	2	1	0.916	0.425
R2	PRI threshold raised to .70	1	1	0.905	0.414
R3	Frequency threshold lowered to n ≥ 1	6	3	0.887	0.571
R4	Alternative anchors: 90th, 50th, 10th percentiles	13	4	0.800	0.660
R5	Alternative outcome: bias-corrected θ	10	3	0.785	0.618
R6	Alternative efficiency model: CCR	11	3	0.910	0.788

Panel B. Configuration stability across outcomes.

Variant	Specification	Models	K	M	K ∨ M
Main	Consistency .80, PRI .50, n ≥ 2, 95/50/5 anchors, raw BCC θ	4	3/3	3/3	3/3
R1	Consistency threshold raised to .85, PRI to .60	1	1/1	1/1	1/1
R2	PRI threshold raised to .70	1	1/1	1/1	1/1
R3	Frequency threshold lowered to n ≥ 1	1	3/3	3/3	3/3
R4	Alternative anchors: 90th, 50th, 10th percentiles	8	3/4	3/4	4/4
R5	Alternative outcome: bias-corrected θ	2	3/3	2/3	3/3
R6	Alternative efficiency model: CCR	1	2/3	2/3	3/3

Note. Rows = truth-table rows coded as producing the outcome. Terms = solution terms in the first equivalent model. K, M, and K ∨ M indicate the numbers of terms containing knowledge sharing, work engagement, and at least one of these conditions, respectively. Across the six variants, knowledge sharing appears in 13 of 15 solution terms and work engagement appears in 12. Every term contains at least one of these conditions.

Figure 14. Robustness of solution consistency and coverage across analytic variants.

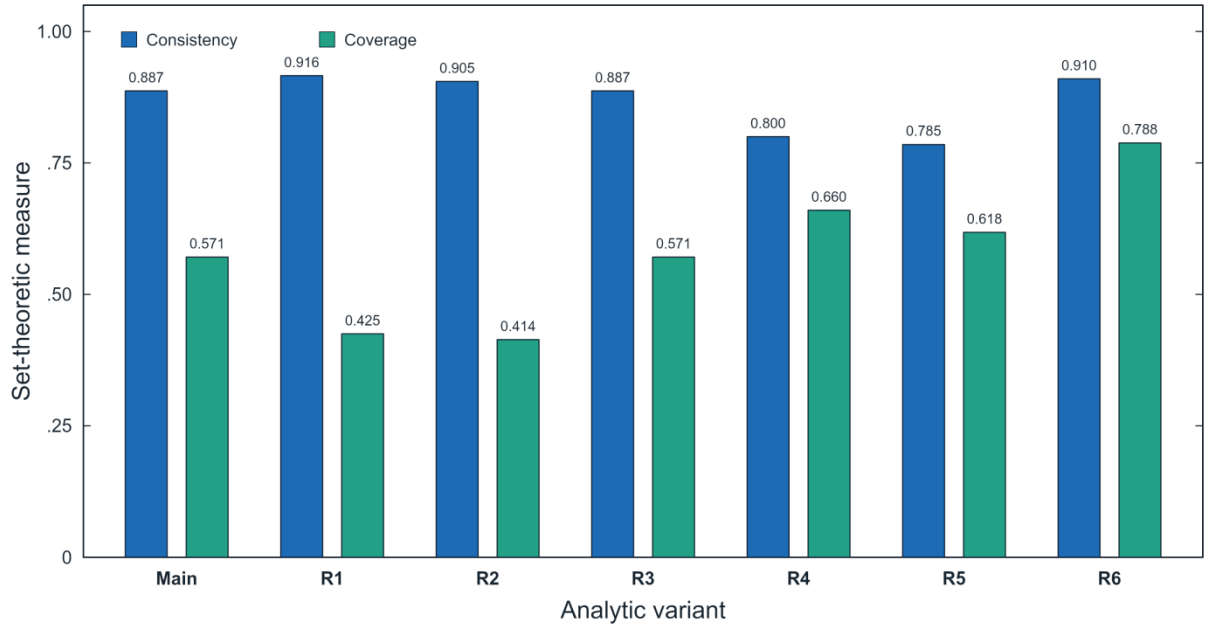
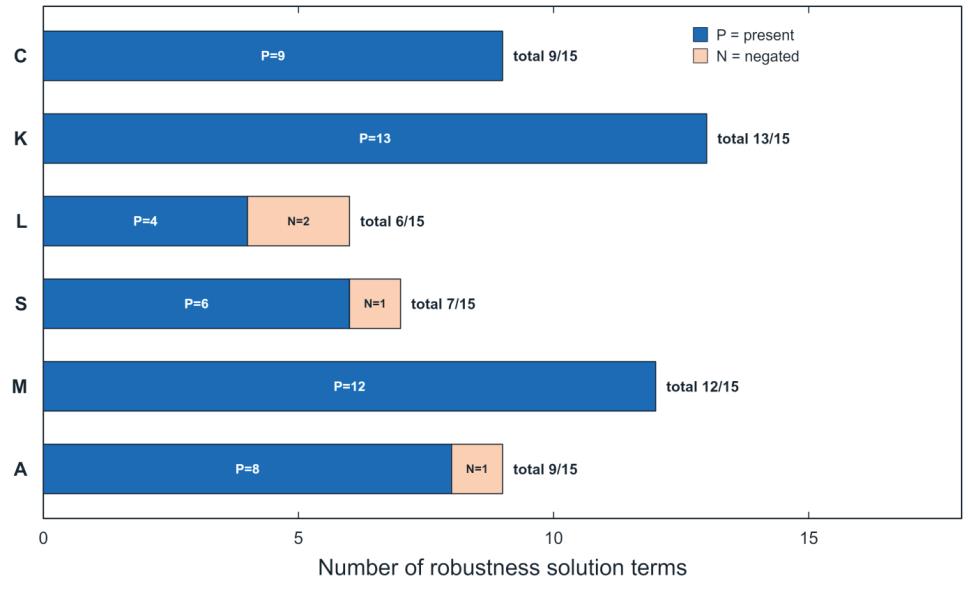


Figure 15. Condition recurrence across robustness solution terms.



The first set of variants changes the thresholds. R1 raises the consistency threshold to .85 and PRI to .60, leaving two truth-table rows and the single term A·M·S·K·C (consistency .916, coverage .425). R2 raises PRI alone to .70, leaving one row and the single term A·M·S·L·K (.905, .414). Both variants increase consistency at the expense of coverage and reduce the solution to one long conjunction. R3 lowers the frequency threshold to $n \geq 1$ and admits a sixth row. It reproduces every term in the main solution (.887, .571) but yields a single model rather than four.

R4 changes the calibration anchors from the 95th/50th/5th percentiles to the 90th/50th/10th percentiles. Under this specification, 13 rows meet the consistency threshold, and the solution expands to four terms: A·M·K + M·~S·K + L·K·C + A·M·S·C (.800, .660). R5 changes the outcome measure by replacing the raw BCC score with the bias-corrected score from the Simar and Wilson procedure. This variant has 10 qualifying rows and three terms: ~A·K·C + M·~L·K + M·K·C (.785, .618). Its consistency is the lowest of the six variants. R6 replaces the BCC efficiency model with CCR, admits 11 rows, and yields L·K·C + A·M·S·C + A·M·~L·K (.910, .788). It has the highest coverage of any variant, and its consistency is exceeded only by R1's .916.

The recurrence of knowledge sharing and work engagement requires precise interpretation. These conditions do not both appear in every solution term of every variant. Across the 15 robustness terms, knowledge sharing appears in 13 and work engagement appears in 12. Each of the 15 contains at least one of the two conditions. Knowledge sharing is absent from A·M·S·C in both R4 and R6. Work engagement is absent from L·K·C in R4, ~A·K·C in R5, and L·K·C in R6. In the main solution, C1, C2, and C3 each contain both conditions. Figure 15 presents the recurrence counts for each condition and its negation.

Negated literals remain in R4 (~S) and R5 (~L, ~A) despite directional expectations specifying the presence of all six conditions. A directional expectation determines which logical remainders are admitted into counterfactual simplification. It does not prevent a negated literal from remaining after minimization when that literal appears in the empirical rows (Duşa, 2019). Intermediate solutions should therefore not be interpreted as if presence had been imposed on the result. R4 also yields eight equivalent models, compared with one for R3 and R6 and two for R5. These differences identify the calibration anchors as the most consequential analytical choice in the chain. Moving the tails by five percentage points changes the number of qualifying rows from 5 to 13, the number of terms from three to four, and the number of equivalent models from four to eight.

5.7 Recovery of the Generating Structure

A known data-generating process allows an assessment that field data cannot provide: the extent to which the analytical chain recovers the true structure. The simulated efficiency scores are generated by max(A·M·S, L·K·C, M·K·~L). Its three paths correspond to propositions P1, P2, and P3. These propositions serve as directional expectations. Figures 16 and 17 assess the seven specifications against this generating structure. Figure 16 shows the recovery status of each generating path, and Figure 17 reports literal-level precision and recall.

Figure 16. Recovery of the known generating structure across analytic variants.

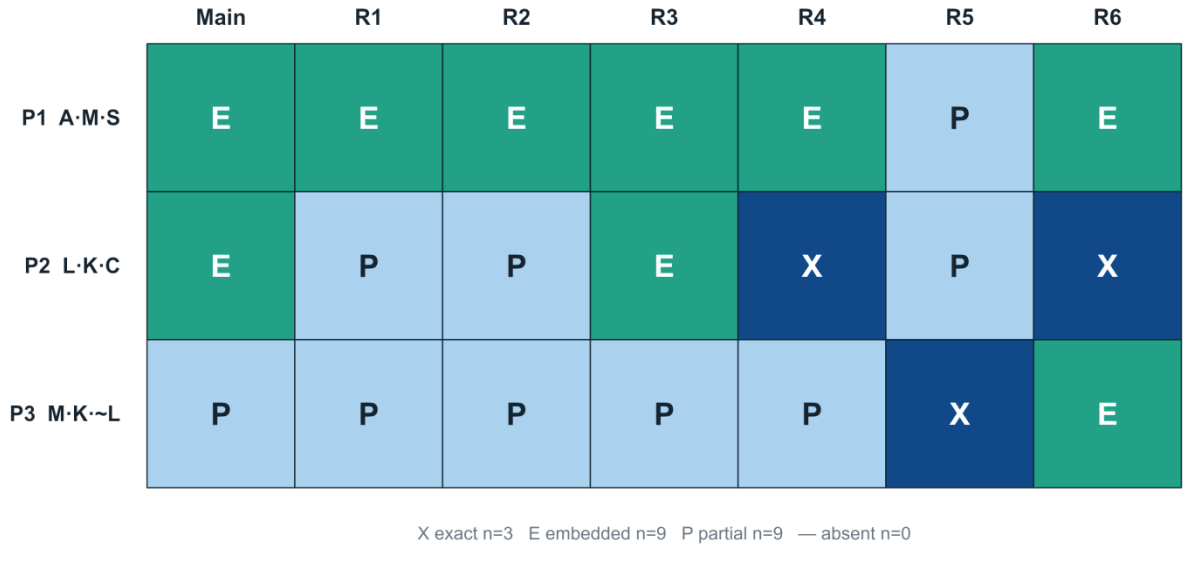
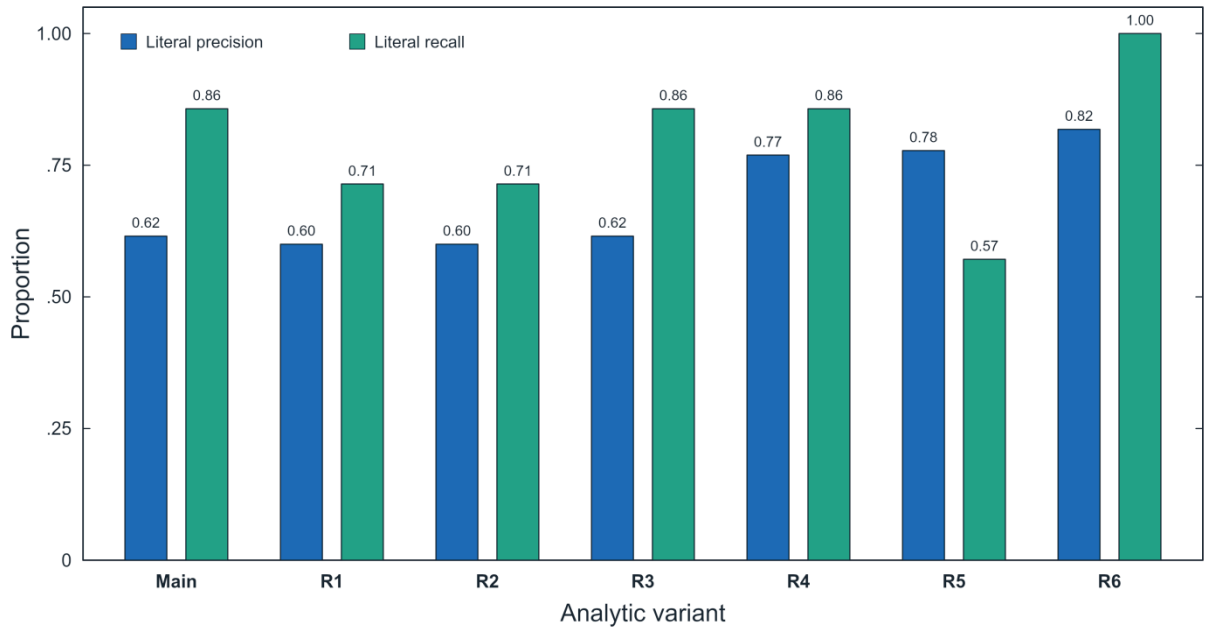


Figure 17. Literal-level precision and recall of recovered solutions.



P1 (A·M·S) is never recovered as an exact term. It is embedded in a longer term under six of the seven specifications and is recovered only partially under R5. In the main solution, it appears within A·M·S·K. P2 (L·K·C) is recovered exactly under R4 and R6. It is embedded in A·M·L·K·C in both the main solution and R3 and is recovered partially under R1, R2, and R5. P3 (M·K·~L) is recovered exactly only under R5, where it appears as M·~L·K. It is embedded in R6's A·M·~L·K and is recovered partially under all other specifications, including the main solution. C2 (M·S·K·C) matches no single generating path and combines literals from two paths.

Figure 17 quantifies this comparison. Precision counts the literals in each recovered term that belong to the generating path with which it overlaps most. Recall counts the distinct generating literals recovered anywhere in the solution. The main specification has precision .62 and recall .86. Tightening the thresholds reduces recall without improving precision: R1 and R2 have precision .60 and recall .71. R3 leaves both values unchanged. Moving the calibration anchors inward increases precision to .77 while leaving recall unchanged in R4. R5 has the highest precision among the BCC-based variants at .78. Its recall of .57 is the lowest among all variants. R6 has precision .82 and recall 1.00.

Knowledge sharing appears in nearly every recovered path. It also appears in the recovered form of P1 despite being absent from P1's true form. This pattern originates before minimization. Among the six conditions, K has the strongest bivariate association with efficiency (Spearman .411), the highest necessity consistency (.757), and the highest necessity coverage (.729). Its calibrated membership is therefore high across most rows meeting the consistency threshold, leaving little basis for minimization to eliminate a literal that rarely contradicts the outcome. A study of real teams could not identify this discrepancy. The additional K in C1 could be interpreted as a substantive finding that knowledge sharing completes the psychological-capital path. The term's consistency, PRI, and coverage would not identify it as a procedural artifact rather than a substantive feature.

R6 recovers the generating structure more accurately than the main specification on every recovery metric reported in Figures 16 and 17. Precision is .82 compared with .62, and recall is 1.00 compared with .86. R6 recovers one generating path exactly and embeds a second, whereas the main specification recovers no path exactly. The specifications differ only in the efficiency model used before calibration: R6 uses CCR and the main specification uses BCC. The main analysis uses BCC because variable returns to scale is a defensible production assumption for teams of different sizes. This choice is not based on superior recovery of the generating structure. For real-world data, this recovery comparison would be unavailable, leaving the modeling assumption as the basis for choosing between the models. The comparison therefore identifies a limitation of the main specification and does not provide a recommendation to calibrate CCR scores.

The results indicate partial recovery of the generating structure. P2 and P3 are recovered exactly only under alternative

specifications. In the main solution, P2 is embedded within C3 and P3 is recovered only partially. P1 is never recovered exactly and appears in the main solution with an additional literal absent from the generating path. The analytical chain identifies the region of the condition space containing the generating paths, but the main solution does not recover those paths exactly.

6. Discussion

6.1 Strengths of the Analytical Chain

No literal is common to all disjuncts of the platform's generating function max(A·M·S, L·K·C, M·K·~L). Accordingly, no condition is expected to meet the .90 necessity convention, and none does: the highest necessity consistency in Table 4 is .757. The sufficiency analysis identifies three configurations, with overall consistency of .887, PRI of .743, and coverage of .571. Their unique coverages are .039, .055, and .052, whereas their raw coverages are approximately .46. These results indicate equifinal paths that overlap within one region of the property space rather than three distinct team types. All 15 terms identified in the robustness analyses include knowledge sharing, work engagement, or both (Table 8). The analytical chain thus preserves the configurational structure of a disjunction of conjunctions rather than reducing it to an additive relationship.

6.2 Choosing the Outcome Measure

Direct calibration is strictly increasing in the raw score. Consequently, two outcome measures that rank the 160 teams identically yield identical truth tables. Calibration therefore retains only the ordering of cases.

The Simar and Wilson correction targets the level of the estimate: it reduces mean θ from 0.551 to 0.489 and produces intervals with a median width of 0.082. However, it reduces agreement between the estimated ranking and the known ranking. The Spearman correlation with the platform's known efficiency falls from .521 for the raw BCC score to .337 for the corrected score. Precision is lowest where the correction is greatest: the median interval width is 0.248 at the frontier, compared with 0.081 elsewhere. One frontier team receives a corrected score of −0.327. This score falls outside the admissible domain (0, 1] and therefore outside the range of any calibration function. These findings support calibrating the raw scores, reporting the corrected scores and their intervals as measures of estimation uncertainty, and including corrected scores in a robustness analysis. R5 yields consistency of .785 and coverage of .618 across two models.

This concern applies to shrinkage and corrected estimates more broadly, not only to Simar and Wilson (2000). Improvements in level accuracy draw on information across units, which changes their ordering. An outcome set based on such estimates inherits those changes.

6.3 Bounded Outcomes and the PRI Threshold

BCC scores are bounded above by 1, with 13 of the 160 teams on the frontier. The outcome's full-membership anchor of 1.000 coincides with this upper bound. This alignment compresses the upper tail of the membership distribution and shifts cases toward the crossover, where simultaneous membership in the outcome and its negation is highest and the PRI penalty is greatest. The resulting reduction in PRI is mechanical.

PRI is the binding criterion. All five qualifying rows in Table 5 have raw consistency of .891 or higher, but three have PRI values between .568 and .600. Row inclusion therefore depends on the PRI threshold. The threshold of .50 reflects this concentration of cases, and increasing it to .70 in R2 leaves one row, one term, and coverage of .414. The choice of PRI threshold should therefore be justified with reference to the outcome distribution (Greckhamer et al., 2018; Schneider & Wagemann, 2012).

6.4 Sources of Potentially Misleading Results

Knowledge sharing has the highest necessity consistency (.757) and the highest Spearman correlation with efficiency (.411), and it appears in all three solution terms. However, it is absent from the generating path A·M·S. C1 = A·M·S·K reproduces that path with a spurious additional literal. This result illustrates how a broadly distributed condition correlated with the outcome can enter a recovered configuration even when it is absent from the corresponding generating path.

The four equivalent intermediate models share C2 and C3 but differ in their first term. The first term is one of four alternatives: A·M·S·K, A·S·~L·K, A·$_{L·K}$C, M·S·~L·K. Their only common literal is K. More permissive calibration increases this ambiguity. R4's 90/50/10 anchors admit 13 rows and yield eight models with consistency of .800. In addition, R4 and R5

include the negated literals ~S, ~L, ~A despite directional expectations favoring the presence of all six conditions.

The distinction between core and peripheral conditions provides no discrimination in this analysis. The intermediate and parsimonious solutions coincide in every model, so the intersection rule (Fiss, 2011) classifies every literal as core.

R6 reveals a further limitation. Calibrating the CCR score yields the best recovery of the generating structure across all reported measures, with precision of .82 compared with .62 and recall of 1.00 compared with .86 (Figures 16 and 17). However, variable returns to scale remains the defensible assumption for teams of different sizes, which is why BCC is used in the main analysis. Better recovery therefore does not necessarily identify the more defensible modeling assumption.

6.5 A Protocol for Field Data

Calibrate raw efficiency scores. Report corrected scores and their intervals to characterize estimation uncertainty, and include corrected scores in a robustness analysis.

Provide a substantive justification for the calibration anchors, and use percentile anchors as a robustness variant rather than the default specification.

Report the treatment of logical remainders and excluded rows in full. In this analysis, there are 16 logical remainders, 8 rows below the frequency threshold, and 5 qualifying rows containing 33 of the 160 teams.

Analyze the negated outcome separately using its own directional expectations, and report it only if it meets the minimum consistency threshold. The solution comprising D1 through D4 has consistency of .706, below the .75 minimum.

Report all equivalent models and designate their intersection as the model-robust core. Before classifying a condition as core, verify that the intermediate and parsimonious solutions differ.

Account for the nested data structure or explicitly acknowledge the inability to do so. The 160 teams are nested within 25 firms, but the truth table treats each team as an independent configuration. Firm-level dependence therefore affects consistency and coverage without an available correction. This issue remains unresolved.

The conditions are simulated draws from a known generating function, and C1 describes the behavior of a minimization algorithm applied to these data. The protocol is transferable to applied research, but the specific paths are not.

7. Limitations

The platform's parameters were specified rather than estimated. Loadings, intraclass correlations, condition correlations, output elasticities, and the form of the generating function were set within ranges reported in published team studies. Table 1 therefore describes simulated item responses, and the findings provide no empirical evidence about R&D teams. The recovery assessment applies to the generating function max(A·M·S, L·K·C, M·K·~L) and to one realization of the generator with seed 20260907. A Monte Carlo analysis across seeds and parameter draws would provide a recovery rate and an error estimate beyond this single case. Six conditions define 64 truth-table rows, which are unevenly populated by the 160 teams. There are 16 logical remainders, 8 rows below the frequency threshold, and 5 rows above the consistency and PRI thresholds containing 33 teams. Minimization yields 4 equivalent intermediate models whose ranking depends on assumptions. Anchors at the 5th, 50th and 95th percentiles define membership relative to these 160 teams. The 90/50/10 variant yields 13 rows and 4 terms, compared with 5 rows and 3 terms in the main specification. Anchors based on an external standard are preferable when such a standard exists (Greckhamer et al., 2018). The radial BCC score contracts all inputs by a common factor without accounting for slack. A slacks-based measure (Cooper et al., 2007) would distinguish teams with the same value of θ and could produce a different truth table. Overall solution consistency for the negated outcome is .706, below the .75 lower bound in set-theoretic practice (Schneider & Wagemann, 2012). Accordingly, D1 through D4 in Table 7 are labeled exploratory, and the difference in term length is descriptive rather than evidence of causal asymmetry. Firm effects enter the generating process, but the truth table treats each of the 160 teams as an independent case. Including firm membership as a condition would address the nesting within 25 firms.

8. Conclusion

This study applied a DEA-to-fsQCA analytical chain to data generated by a known process. The analysis comprised 160 simulated teams with mean raw BCC efficiency of 0.551. It identified three configurations associated with high efficiency.

The solution has consistency of .887 and coverage of .571. The chain reproduced the broad structure of the generating function: no necessary condition, conjunctions of four and five conditions above the consistency threshold, and solution consistency above the .80 convention. Two lessons are relevant to applied research. First, calibrate the raw efficiency score rather than the bias-corrected score. Calibration depends on relative position, and the Simar and Wilson correction improved the point estimate but reduced agreement with the true ordering (ρ = .52 raw against .34 corrected). The corrected estimate and its interval should therefore be reported as measures of estimation uncertainty. Second, justify the PRI threshold when the outcome set is bounded. If the full-membership anchor coincides with the theoretical maximum, memberships concentrate at the frontier and PRI is reduced near the crossover. Recovery remains partial: one generating path is recovered with a spurious condition, another appears only as a subset of a larger term, and the third is recovered only under alternative specifications. The study provides a methodological template for applying the same analytical chain to field data.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Appelbaum, E., Bailey, T., Berg, P., & Kalleberg, A. L. (2000). Manufacturing advantage: Why high-performance work systems pay off. ILR Press.

[2] Avey, J. B., Reichard, R. J., Luthans, F., & Mhatre, K. H. (2011). Meta-analysis of the impact of positive psychological capital on employee attitudes, behaviors, and performance. Human Resource Development Quarterly, 22(2), 127-152. https://doi.org/10.1002/hrdq.20070

[3] Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30(9), 1078-1092. https://doi.org/10.1287/mnsc.30.9.1078

[4] Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions (pp. 349-381). Jossey-Bass.

[5] Blumberg, M., & Pringle, C. D. (1982). The missing opportunity in organizational research: Some implications for a theory of work performance. Academy of Management Review, 7(4), 560-569. https://doi.org/10.5465/amr.1982.4285240

[6] Carmeli, A., Reiter-Palmon, R., & Ziv, E. (2010). Inclusive leadership and employee involvement in creative tasks in the workplace: The mediating role of psychological safety. Creativity Research Journal, 22(3), 250-260. https://doi.org/10.1080/10400419.2010.504654

[7] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2(6), 429-444. https://doi.org/10.1016/0377-2217(78)90138-8

[8] Cooper, W. W., Seiford, L. M., & Tone, K. (2007). Data envelopment analysis: A comprehensive text with models, applications, references and DEA-Solver software (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-45283-8

[9] Du, Y., & Jia, L. (2017). Zutai shijiao yu dingxing bijiao fenxi (QCA): Guanlixue yanjiu de yi tiao xin daolu [A configurational perspective and qualitative comparative analysis (QCA): A new path for management research]. Guanli Shijie [Journal of Management World], 33(6), 155-167. https://doi.org/10.19744/j.cnki.11-1235/f.2017.06.012 (in Chinese)

[10] Duşa, A. (2019). QCA with R: A comprehensive resource. Springer. https://doi.org/10.1007/978-3-319-75668-4

[11] Edmondson, A. (1999). Psychological safety and learning behavior in work teams. Administrative Science Quarterly, 44(2), 350-383. https://doi.org/10.2307/2666999

[12] Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society: Series A (General), 120(3), 253-281. https://doi.org/10.2307/2343100

[13] Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. Academy of Management Journal, 54(2), 393-420. https://doi.org/10.5465/amj.2011.60263120

[14] Greckhamer, T., Furnari, S., Fiss, P. C., & Aguilera, R. V. (2018). Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. Strategic Organization, 16(4), 482-495. https://doi.org/10.1177/1476127018786487

[15] James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. Journal of Applied Psychology, 69(1), 85-98. https://doi.org/10.1037/0021-9010.69.1.85

[16] Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. Academy of Management Journal, 33(4), 692-724. https://doi.org/10.5465/256287

[17] Luthans, F., Avolio, B. J., Avey, J. B., & Norman, S. M. (2007). Positive psychological capital: Measurement and relationship with performance and satisfaction. Personnel Psychology, 60(3), 541-572. https://doi.org/10.1111/j.1744-6570.2007.00083.x

[18] Misangyi, V. F., & Acharya, A. G. (2014). Substitutes or complements? A configurational examination of corporate governance mechanisms. Academy of Management Journal, 57(6), 1681-1705. https://doi.org/10.5465/amj.2012.0728

[19] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

[20] Ragin, C. C. (2000). Fuzzy-set social science. University of Chicago Press.

[21] Ragin, C. C. (2008). Redesigning social inquiry: Fuzzy sets and beyond. University of Chicago Press. https://doi.org/10.7208/chicago/9780226702797.001.0001

[22] Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. Educational and Psychological Measurement, 66(4), 701-716. https://doi.org/10.1177/0013164405282471

[23] Schneider, C. Q., & Wagemann, C. (2012). Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis. Cambridge University Press. https://doi.org/10.1017/CBO9781139004244

[24] Scott, S. G., & Bruce, R. A. (1994). Determinants of innovative behavior: A path model of individual innovation in the workplace. Academy of Management Journal, 37(3), 580-607. https://doi.org/10.5465/256701

[25] Siemsen, E., Roth, A. V., & Balasubramanian, S. (2008). How motivation, opportunity, and ability drive knowledge sharing: The constraining-factor model. Journal of Operations Management, 26(3), 426-445. https://doi.org/10.1016/j.jom.2007.09.001

[26] Simar, L., & Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. Journal of Applied Statistics, 27(6), 779-802. https://doi.org/10.1080/02664760050081951

[27] Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 136(1), 31-64. https://doi.org/10.1016/j.jeconom.2005.07.009

[28] Thiem, A., & Duşa, A. (2013). QCA: A package for qualitative comparative analysis. The R Journal, 5(1), 87-97. https://doi.org/10.32614/RJ-2013-009

[29] Van den Hooff, B., & de Ridder, J. A. (2004). Knowledge sharing in context: The influence of organizational commitment, communication climate and CMC use on knowledge sharing. Journal of Knowledge Management, 8(6), 117-130. https://doi.org/10.1108/13673270410567675