Journal of Mechanical, Civil and Industrial Engineering
ISSN: 2709-0809
DOI: 10.32996/jmcie
Journal Homepage: www.al-kindipublisher.com/index.php/jmcie

JMCIE
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

A Risk-Based Safety Resilience Framework for Brownfield Restoration Projects

Shajahan Abdul Salam
Principal HSE Specialist, Kellogg Brown & Root-AMCDE, Al Khobar, Saudi Arabia
Corresponding Author: Shajahan Abdul Salam**, E-mail**: shajahan.abdulsalam@kbr.com

| ABSTRACT

Brownfield restoration combines construction hazards with residual process risk, temporary systems, changing plant states and simultaneous operations. This framework-development paper proposes the Restoration Safety Resilience Index (RSRI) as a structure for organizing evidence, not as a validated measure of safety performance. Four resilience potentials—Anticipation, Monitoring, Response and Learning—are linked to Interface Coordination and Operational Readiness. An anonymized Saudi Arabian restoration case provides an author's retrospective illustration of how reported operational records might be mapped to these dimensions. A qualitative evidence-quality assessment distinguishes completeness, contemporaneity, independent verification and recurrence information. Incomplete item-level provenance and the absence of independent coding prevent calculation of a defensible empirical project score; all numerical RSRI examples are therefore confined to a worked-example appendix under author-proposed assumptions. Permit Interaction Load is presented separately as a proposed future metric and was not operationalized in the case. Comments from nine practitioners inform wording refinement only. The contribution is a transparent evidence-mapping proposal, a set of source-quality constraints and a testable validation pathway. Independent double-coding, expert calibration and prospective multi-project testing remain prerequisites for measurement claims or score-based assurance decisions.

| KEYWORDS

Resilience engineering; Safety-II; brownfield construction; restoration; process safety; SIMOPS; permit to work; leading indicators; construction safety; RSRI.

| ARTICLE INFORMATION

ACCEPTED: 25 August 2026 **PUBLISHED:** 22 September 2026 **DOI:** 10.32996/jmcie.2026.7.5.5

1. Introduction

Brownfield restoration inherits hazards from existing plant, damaged structures, residual process materials and temporary systems. Risk changes as isolations move, equipment is opened, materials are disturbed and areas are handed back. Concurrent tasks can introduce ignition, atmospheric contamination, line-of-fire exposure or restricted emergency access even when each task is acceptable in isolation.

Traditional occupational safety systems remain essential in this setting. Permit-to-work (PTW), lockout/tagout and positive isolation, gas testing, confined-space entry control, lifting plans, scaffolding inspection, demolition engineering, personal protective equipment, competency management, and emergency arrangements provide the fundamental barriers. However, these controls are often administered as task-level requirements. The core restoration challenge is therefore not only whether each task possesses its own controls, but whether the entire work system continues to operate safely when tasks, people, equipment, and plant states interact. This system-level question is closely aligned with resilience engineering, which focuses on the ability to respond, monitor, learn, and anticipate under varying conditions (Hollnagel, 2014, 2017; Hollnagel et al., 2011; Provan et al., 2020).

Construction safety research supports proactive leading indicators alongside lagging measures, including management commitment, stakeholder engagement, competence, field verification and learning (Hinze et al., 2013; Shaikh et al., 2020; Xu et al., 2021, 2023). Resilience-oriented safety-climate research links learning, reporting and anticipation to construction safety (Chen et al., 2018). Mollo (2026) examines construction management through a Safety-II perspective, emphasizing practices that support adaptation under variable conditions. RSRI builds on this established work; it does not claim to introduce resilience or evidence-based leading indicators to construction.

For this study, safety resilience is operationally defined as the evidenced capability of the restoration work system to anticipate changing conditions, detect deterioration, respond to deviations, learn from experience, coordinate interacting work, and maintain readiness without violating critical-control boundaries. Resilience is therefore not equated with the absence of injuries or with a high volume of compliance activity.

Routine HSE records are often reported separately, producing activity counts without showing whether controls remain effective as work changes. RSRI integrates these records into a project-level evidence map of adaptive safety capability.

The proposed contribution is a restoration-specific organization of six capability dimensions, explicit evidence-quality constraints and critical-control overrides. PIL is a separate proposal for future permit-interaction screening. Neither the index nor PIL is validated or demonstrated as an operational risk measure in this study.

1.1 Research aim and objectives

The aim is to specify a restoration-focused framework that can be tested in future research. Objectives are to characterize changing restoration hazards; organize routine evidence using resilience and PTW/SIMOPS concepts; define transparent scoring and evidence-quality rules; illustrate the proposed mapping with anonymized operational summaries; and specify the reproducibility, calibration and prospective testing required before numerical use.

1.2 Research questions

RQ1: Which resilience-related capabilities are relevant to changing restoration work?

RQ2: Which routine operational records could support assessment, and where does their quality prevent scoring?

RQ3: How can critical-control failures and evidence limitations remain visible in a composite framework?

RQ4: What validation steps are needed before numerical RSRI or PIL use?

2. Literature Review and Theoretical Basis

2.1 From Safety-I to resilient performance

Conventional safety management, often described as Safety-I, is primarily concerned with preventing unwanted outcomes by identifying hazards, specifying controls, and correcting deviations. This remains indispensable for construction and process operations. Resilience engineering does not replace this logic; it adds a complementary question: how does a socio-technical system continue to succeed when actual conditions differ from plans? Hollnagel describes four systemic potentials that support resilient performance: the ability to respond to disturbances and opportunities, monitor what may affect performance in the near term, learn from experience, and anticipate future developments (Hollnagel, 2014, 2017; Hollnagel et al., 2011). Provan et al. further distinguish centralized control from guided adaptability and argue that safety professionals can create value by developing foresight about changing risk and enabling safe adaptation before harm occurs (Provan et al., 2020).

Figure 1 translates the four resilience potentials into restoration practice and shows the two enabling capabilities. Monitoring, response and learning must continue after permit issue because adjacent work, plant state and temporary controls can change during execution.

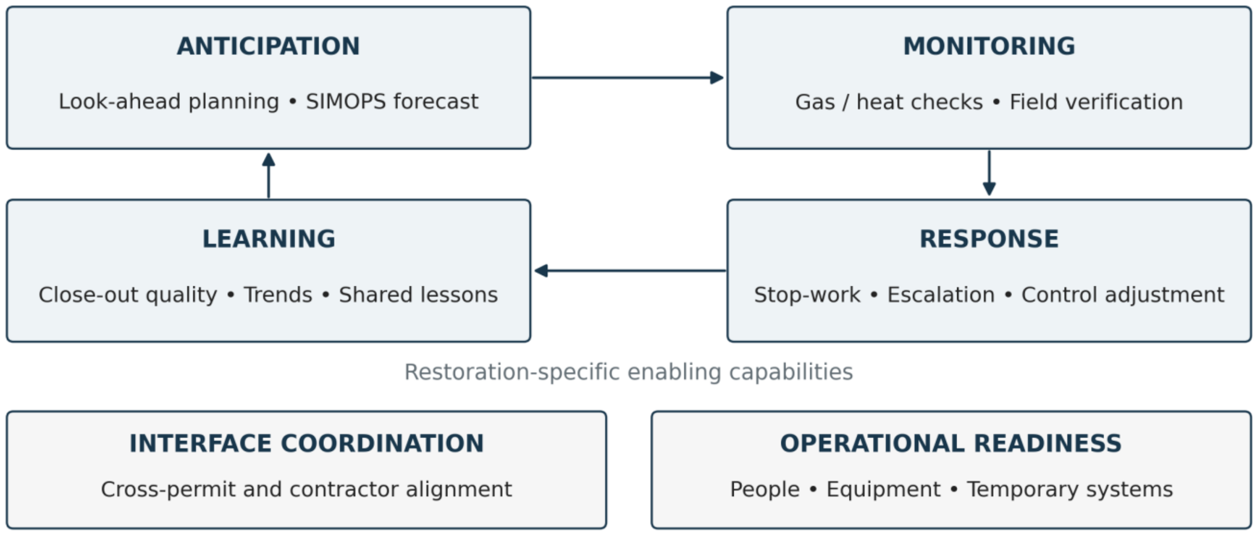

Figure 1. Core resilience potentials and restoration-specific enabling capabilities. Author's synthesis based on Hollnagel (2014, 2017) and Provan et al. (2020).

2.2 Construction leading indicators

Lagging indicators such as recordable injury rates are useful for governance and benchmarking but are weak diagnostic tools for changing conditions. A low injury count can arise from effective controls, low exposure, chance, underreporting, or insufficient time at risk. Xu et al. identified sixteen leading indicators in construction and argued for an ecosystem perspective that considers individual, project, and organizational levels (Xu et al., 2021). Subsequent work found that organizational commitment, client engagement, contractor engagement, training and orientation, safety climate, and competence are among the most critical proactive indicators (Xu et al., 2023). Reviews of safety performance indicators similarly emphasize management commitment, site audits, near-miss reporting, orientation, process controls, and technology (Shaikh et al., 2020).

The value of leading indicators is not simply that they occur before an incident. A useful indicator must have a plausible connection to the control system, be measurable with acceptable burden, be sensitive to deterioration or improvement, and trigger a management response. Therefore, the RSRI developed in this paper emphasizes evidence quality and control effectiveness rather than counting activities indiscriminately. For example, the number of toolbox talks is less meaningful than whether task changes were communicated and understood; the number of observations is less meaningful than whether repeat conditions decline and actions are closed effectively.

A recent comprehensive review of construction safety leading indicators confirms both the value and the measurement challenge of proactive indicators: definitions and implementation practices continue to evolve, and direct relationships with safety outcomes remain difficult to establish because performance is shaped by interacting organizational and project factors (Golabchi et al., 2024). This supports the present study's emphasis on traceable evidence, trend direction, and management response rather than treating any single activity count as a surrogate for safety.

2.3 Safety climate, learning and adaptive capacity

Construction safety climate studies show that management commitment, supervisory role, rules and procedures, individual responsibility, and training are associated with safety performance (Alruqi et al., 2018). Chen et al. extended the concept by adding resilience-related dimensions such as learning, reporting, anticipation, and awareness (Chen et al., 2018). These constructs connect directly with restoration execution. A project may possess technically strong procedures but remain brittle if information is not shared across shifts, workers do not report emerging hazards, supervision cannot reconfigure controls quickly, or lessons remain isolated within one contractor. Conversely, a project that routinely detects and corrects weak signals can maintain safe performance despite operational variability.

2.4 PTW, SIMOPS and process-safety interfaces

PTW systems are intended to coordinate hazardous work, including interactions between tasks; their purpose extends beyond individual permit compliance. HSG250 explicitly addresses permit interaction, suspension, shift handover and hand-back (Health and Safety Executive [HSE], 2005). Foundational PTW work also emphasizes responsibility, isolation and communication (Booth & Butler, 1992), while Kim and Jung (2011) integrate JSA with permit structures. The concern addressed here is whether these coordination requirements remain visible and verified as restoration conditions change.

SIMOPS risk assessment addresses this issue by identifying conflicts between concurrent activities and updating mitigations as work progresses (Marucco, 2016). Research in other high-hazard domains has emphasized the time-dependent nature of SIMOPS and the limits of static risk models (Fan et al., 2021, 2023). This supports the use of a resilience framework: the risk created by interactions is inherently dynamic, so monitoring and anticipation must continue after initial authorization. The RSRI therefore treats interface coordination as a distinct dimension rather than assuming that individual permit compliance guarantees system safety.

Digital PTW records also create an opportunity to move beyond static compliance reporting. Process-mining research using PTW event logs in an Australian energy company showed that timestamped workflow data can reveal process complexity, conformance deviations, bottlenecks, and organizational patterns that are difficult to see in conventional summaries (Pika et al., 2021). For the RSRI, this is important because the proposed PIL and several leading indicators can be computed from the same fields already captured by mature e-PTW systems: permit identity, location, time window, isolation, status, owner, and workflow timestamps.

2.5 Confined space, atmospheric and restoration-specific hazards

Confined spaces illustrate the combined physical and atmospheric hazards of restoration. Recent reviews of construction confined-space work identify atmospheric hazards, physical hazards, procedural deficiencies, and management failures as recurring contributors to severe events (Arifin et al., 2023; Eustáquio et al., 2024). In hydrocarbon restoration, this can include oxygen deficiency or enrichment, flammable atmospheres, hydrogen sulfide, benzene and other VOCs, residual pressure, pyrophoric deposits, contaminated sludge, temperature, access limitations, difficult rescue geometry, and interaction with adjacent work. Similar dynamics exist in demolition, decontamination, line opening, and critical lifting, where the hazard state can change as material is disturbed or the structural configuration changes.

2.6 Process-safety metrics and barrier health

Process-safety guidance combines leading and lagging information to examine barrier health before losses occur. CCPS provides a process-safety metrics guide (Center for Chemical Process Safety [CCPS], 2022), and API's announcement describes the third edition of RP 754 and its role in continuous assessment (American Petroleum Institute [API], n.d.). RSRI draws on the principle of reviewing control evidence rather than inferring effectiveness from injury counts.

Implementation is not purely a technical problem. Recent construction research reports that organizational capability, leadership, workforce participation, data quality, and implementation barriers can limit the effective use of leading indicators even when the measures themselves are well designed (Adekunle et al., 2024). Accordingly, the framework treats governance and evidence quality as part of resilience rather than assuming that dashboard availability alone improves safety.

2.7 Research gap

Related work already develops leading indicators, resilience-based performance-measurement frameworks and analyses of operational records (Guo & Yiu, 2016; Peñaloza et al., 2021; Pika et al., 2021; Xu et al., 2021, 2023). The narrower contribution proposed here is their integration around changing restoration interfaces, readiness and explicit limits on evidence use. Table 1 compares these approaches without claiming that they lack capabilities not examined in this paper. Table 2 summarizes restoration conditions and literature-informed failure mechanisms; it does not report observed case incidents.

Table 1. Positioning of RSRI relative to related approaches

Approach	Established focus	Relationship to this proposal
Guo and Yiu (2016); Xu et al. (2021, 2023)	Construction leading indicators and their implementation	Provides the indicator foundation; RSRI proposes a restoration-specific evidence organization, not a replacement indicator theory.
Peñaloza et al. (2021)	Resilience-based assessment of safety performance measurement systems	Direct conceptual antecedent. RSRI narrows attention to changing restoration interfaces and source-quality restrictions.
Pika et al. (2021)	Process mining of safety-related operational event data	Shows that operational records can already support analysis. PIL would require a separately specified permit-pair dataset and validation.
HSE (2005); Marucco (2016)	PTW interaction, coordination and simultaneous operations	Establishes the coordination basis. RSRI does not claim that permit-interaction control is new.
Present framework	Six proposed restoration capability dimensions with evidence constraints	No empirical composite score, operational PIL series, reliability estimate or demonstrated superiority.

Table 2. Restoration conditions, potential failure mechanisms and relevant capabilities

Restoration characteristic	*Dynamic risk implication*	*Required resilient capability*	*Potential failure mechanism*
Live or partially live plant interfaces	*Plant state and construction work can interact; isolation boundaries may change*	*Anticipation, monitoring, interface coordination*	*Loss of isolation or unclear risk ownership (HSE, 2005)*
Residual hydrocarbons / toxic contaminants	*Atmospheric condition can change after opening, cleaning, disturbance or adjacent work*	*Continuous monitoring, rapid response*	*Atmospheric deterioration during changing work (Arifin et al., 2023)*
Fire-damaged or degraded structures	*Load paths and access conditions may be uncertain and change during demolition*	*Engineering readiness, change recognition, exclusion control*	*Uncertain load path; conceptual restoration scenario*
High permit density	*Individually acceptable tasks can conflict when spatially or temporally combined*	*SIMOPS mapping, permit coordination*	*Unrecognized interaction or incompatible work (HSE, 2005)*
Multiple contractors and shifts	*Information can degrade at organizational and shift boundaries*	*Structured handover, communication, learning*	*Incomplete handover or cross-team communication (HSE, 2005)*
Temporary ventilation / scaffolds / access	*Temporary systems become critical safety barriers*	*Barrier verification, inspection and contingency readiness*	*Degraded temporary barrier; conceptual scenario*
Night work and heat exposure	*Human performance and environmental conditions vary over time*	*Fatigue / heat monitoring, work-rest adaptation*	*Delayed detection or control response; conceptual scenario*
Progressive handover / start-up interface	*Risk ownership and work boundaries move between teams*	*Governance clarity, readiness reviews, transfer control*	*Premature or incomplete hand-back (HSE, 2005)*

3. Methodology

3.1 Research design

This is a framework-development study with an author's retrospective illustration based on anonymized operational summaries from a hydrocarbon restoration project in Saudi Arabia. The author had an operational HSE role. No independent researcher coded a standardized source-record package, and the summaries do not provide a complete item-by-item audit trail. The case is therefore used to identify potential evidence sources and their limitations, not to estimate an empirical RSRI, demonstrate reproducibility or evaluate intervention effectiveness. The single-case setting also limits transferability (Yin, 2018).

Three evidence types are kept separate: reported case summaries, author-proposed scoring assumptions and practitioner wording comments. The numerical illustration in Appendix C preserves an author-assigned input profile solely to explain arithmetic. It is not reconstructed from item-level case codes. The nine-practitioner exercise did not involve independent coding of the same evidence and supplies no reliability evidence for the case mapping.

Reported records were considered for qualitative mapping when they concerned restoration activities and had an identifiable operational meaning. Missing dates, denominators, source identifiers or recurrence histories restrict interpretation. Figure 2 shows the proposed screening and coding pathway; it is not a flow diagram of a completed, independently verified coding exercise. No source-document counts or missing values have been invented.

3.2 Case-study scope

The restoration scope included isolation and decontamination, removal of damaged equipment, demolition and scrap handling, line cutting and blinding, scaffolding, lifting, hot work, confined-space activities, cleaning, temporary ventilation, waste handling, inspection, and progressive handover. The work was executed in a live brownfield context with simultaneous construction and operational interfaces. Key hazards included residual hydrocarbons, hydrogen sulfide, benzene and VOC exposure, oxygen deficiency, pyrophoric iron sulfide, contaminated sludge, unstable or fire-affected structures, ignition sources, line-of-fire exposure, dropped objects, temporary electrical systems, environmental heat, and night-work human-factor risks.

Table 3. As-reported case summaries and limits on their analytical use

Evidence item	***As-reported value / practice***	***How used in analysis***
Recorded work-hours	*1,636,524 accumulated work-hours as of 20 Jun in the documented performance series; a later supplied project-period figure was approximately 2.7 million work-hours*	*Longitudinal exposure context; not treated as proof of resilience or causality*
Recordable outcomes	*No reported injury / recordable incident in the supplied period*	*Lagging outcome; interpreted cautiously*
Permit load	*54 active permits at an undated representative snapshot, including 14 client-controlled permits*	*Reported permit context only; no observed PIL series*
Night-shift observations	*567 reported night-shift safety-observation close-outs; no linked item-level verification or recurrence series supplied*	*Activity volume only; insufficient to establish response or learning effectiveness*
Worker orientation	*113 workers documented in an initial orientation dataset*	*Readiness / competence evidence*
HSE coverage	*24/7 monitoring with shift handover; HSE officers increased from 25 (15 Jun) to 26 (18 Jun) and 28 (20 Jun) as manpower increased*	*Monitoring and communication evidence*
SIMOPS	*Three-day look-ahead / coordination and risk review*	*Anticipation and interface coordination evidence*
Risk assessment	*Dynamic HIRA, JSA / pre-task risk assessment, start-card practices*	*Anticipation and response evidence*

Evidence item	*As-reported value / practice*	*How used in analysis*
Atmospheric controls	*Gas testing and continuous/task-specific H2S/LEL/O2 monitoring for relevant work*	*Monitoring and critical-barrier evidence*
Confined spaces	*Ventilation, rescue readiness, communications and atmospheric monitoring requirements*	*Readiness and response evidence*
Heat stress	*Seven reported heat-stress logs; approximately 30-minute monitoring where applied; reported work/rest adaptations and no reported heat illness*	*Potential M3/R1 evidence; no complete measurement-to-action-to-recovery linkage*
Scaffolding	*Engineered / inspected scaffolds and green-tag readiness control*	*Operational Readiness evidence*
Emergency preparedness	*Approximately 80% mock-drill participation and evacuation under 10 minutes reported; denominator and standardized timing series not supplied*	*Descriptive response/readiness claim, not validated response performance*
Workforce scale-up	*Manpower increased from 1,016 (15 Jun) to 1,209 (18 Jun) and 1,245 (20 Jun); daily work-hours increased from 12,192 to 14,508 and 14,940*	*Exposure normalization and resource scaling*
Coordination effectiveness	*Approximately 95% coordination attendance and permit turnaround below two hours reported; denominators and full timestamps unavailable*	*Descriptive coordination claim; does not establish interaction-control effectiveness*
Critical activity assurance	*100% monitoring of identified critical activities reported; 11 activities in a 9 June snapshot; completeness not independently verified*	*Monitoring-coverage claim requiring denominator and source-level verification*
Learning / repeat findings	*Approximately 50% reduction in repeat findings reported; baseline, follow-up denominator and linked recurrence history unavailable*	*Unverified improvement claim; excluded from numerical scoring*
Hot-work resilience	*More than 80% of rerun hot-work hours reported as completed under habitat arrangements; denominator and verification scope unavailable*	*Descriptive account only; not proof of adaptive-control effectiveness*

Table 4. Timing and analytical status of reported case evidence

Period / snapshot	*Observed or reported evidence*	*Analytical status*	*Use in manuscript*
9 Jun snapshot	*11 high-risk activities tracked*	*Observed project snapshot*	*Monitoring context only*
15 Jun	*1,016 manpower; 12,192 daily work-hours; 25 HSE officers*	*Observed project series*	*Exposure / assurance-load context*
18 Jun	*1,209 manpower; 14,508 daily work-hours; 26 HSE officers*	*Observed project series*	*Exposure / assurance-load context*
20 Jun	*1,245 manpower; 14,940 daily work-hours; 1,636,524 cumulative work-hours; 28 HSE officers*	*Observed project series*	*Exposure / assurance-load context*
Representative permit snapshot	*54 active permits; 14 client-controlled*	*Observed snapshot*	*Interface-density context; not a PIL score*

Period / snapshot	***Observed or reported evidence***	***Analytical status***	***Use in manuscript***
Referenced project datasets	*567 night-shift observation close-outs; 113 worker orientations; seven heat-stress logs*	*Observed / reported evidence*	*Dimension evidence where timing and meaning were clear*
Later project milestone	*Approximately 2.7 million cumulative work-hours*	*Later reported milestone*	*Exposure context only; not synchronized with June snapshots*

3.3 Data integrity and missing-data handling

The summaries span different periods. The dated 15–20 June workforce series is distinct from the representative permit snapshot and the later cumulative-hours milestone; no common assessment window has been established across them. No missing timestamp, denominator or record identifier is imputed. These limitations prevent a synchronized quantitative capability assessment.

Tables 5 and 7 assess what the available summaries demonstrate, not the quality of confidential underlying records that were unavailable for independent inspection. Each quality domain is labelled Full (F), Partial (P) or Not demonstrated (ND), using the explicit rubric below. ND means missing support in the supplied material, not proof that a control or verification activity did not exist.

Table 5. Proposed evidence-quality rubric applied to the available summaries

Quality domain	**Full (F)**	**Partial (P)**	**Not demonstrated (ND)**
Completeness (C)	Required source IDs, population/coverage and denominators available	Some context or counts; coverage or denominator incomplete	No defensible source-level description
Contemporaneity (T)	Evidence belongs to the same defined review window	Some dates or context, but a mixed or undefined assessment window	Period or alignment cannot be established
Independent verification (V)	Independent reviewer corroborated control status against a source	Verification described but independence or scope incomplete	No independent corroboration demonstrated
Recurrence / feedback (R)	Linked baseline, corrective action and repeated follow-up evidence	Feedback described but linkages or follow-up incomplete	No linked recurrence or feedback series available

Derived values were recalculated from the reported source numbers and rounded only at the presentation stage. Percentage change was calculated as (later value - earlier value) / earlier value × 100. Permit share was calculated as client-controlled permits / total active permits × 100, and assurance load was described as manpower / HSE officers. These are descriptive exposure and governance indicators; they are not interpreted as causal predictors of injury performance.

Quality ratings are not averaged into an evidence-confidence score. The previous global "Moderate" label is withdrawn because no calibrated rule supported it. For numerical assessment, source identity, applicable period, expected-control coverage and item-code rationale must be recoverable. Until that minimum is met, the assessment is "not assessable from the available summaries," even if operational practices appear plausible.

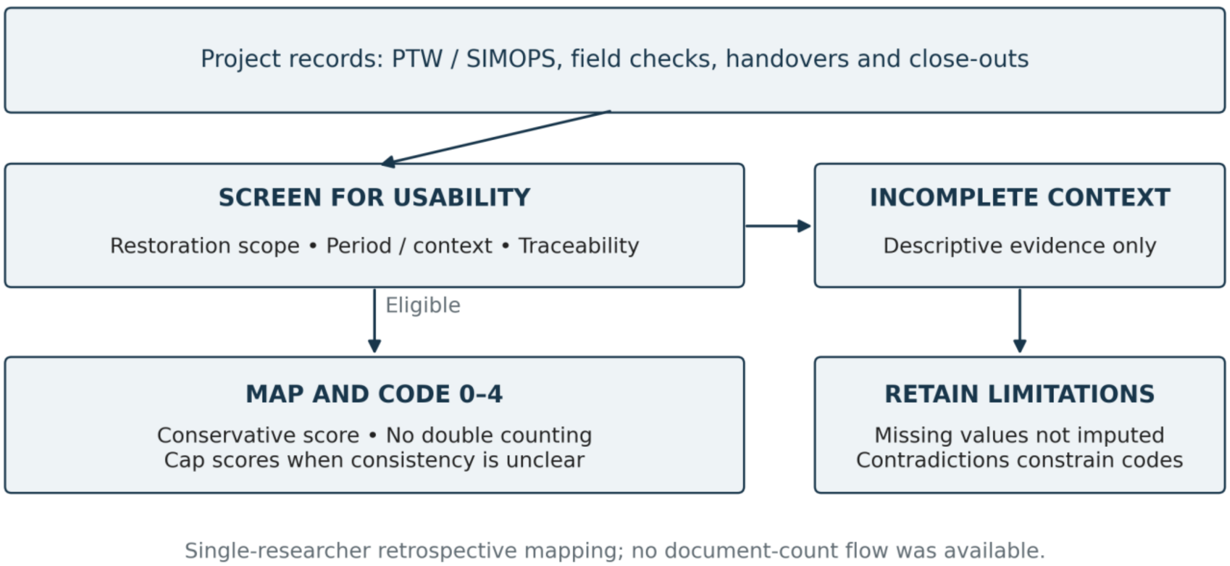

Figure 2. Proposed evidence screening and conservative coding pathway; not a completed reliability study.

3.4 Proposed item coding and evidence constraints

Appendix B defines an ordinal 0–4 item scale: 0 requires demonstrated absence or contradiction; 1 denotes ad hoc evidence; 2 denotes a defined control without adequate consistency evidence; 3 requires repeated implementation and field verification; and 4 additionally requires documented feedback or adaptation. Unavailable evidence is marked NE (not evaluable), not 0. No code is assigned merely because an activity count is high. Table 7 identifies specific item constraints and the additional records needed to apply these rules.

For dimension d, D_d = (mean of applicable 0–4 item codes / 4) × 100. Equivalently, $D_d = [\Sigma c_{dk} / (4n_d)] \times 100$, where c_{dk} is an admissible item code and n_d is the number of scored applicable items. Report the total expected applicable items, exclusions, NE items and source coverage alongside every dimension. If $n_d = 0$, D_d is undefined. Excluding NE items must not silently inflate a score; a composite is withheld until a pre-specified coverage rule is satisfied. No such coverage rule has been validated in this study.

3.5 Proposed aggregation structure

The generic weighted index is:

$RSRI = \Sigma\, w_d D_d$, for $d \in \{A, M, R, L, I, O\}$, with $w_d \geq 0$ and $\Sigma\, w_d = 1$.

A, M, R and L denote the four resilience potentials; I and O denote restoration-specific enabling capabilities. All dimension inputs are on a 0–100 scale. The proposed dimensions and candidate evidence are shown in Table 6. Neither the dimension structure nor its weighting is empirically established here.

Numerical weights are not justified by practitioner voting or an expert elicitation: neither was undertaken. The original author-proposed weights and their arithmetic sensitivity are confined to Appendix C. They are illustrative assumptions rather than recommended coefficients. Future expert review must examine content overlap, the adequacy of Learning and Readiness representation, and the consequences of alternative weighting choices.

Table 6. Proposed RSRI dimensions and candidate evidence

Dimension	*What it asks*	*Example evidence / metric*
Anticipation (A)	*Can the project foresee changing hazards and interactions?*	*3-day look-ahead; SIMOPS forecast; dynamic HIRA; planned isolation changes; weather/heat forecast; critical-lift readiness*
Monitoring (M)	*Can the project detect deterioration and weak signals early?*	*Field verification; gas monitoring; heat/lux checks; scaffold inspection; permit audit; repeat observation rate*
Response (R)	*Can the project act quickly and proportionately when conditions change?*	*Stop-work; escalation time; evacuation/drill performance; corrective action; control reconfiguration*
Learning (L)	*Does information from work change future work?*	*Observation close-out quality; trend reviews; lessons learned; shift handover; repeat finding reduction*
Interface Coordination (I)	*Are cross-permit, cross-contractor and client interfaces actively controlled?*	*SIMOPS matrix; permit conflict review; exclusion-zone coordination; daily stakeholder meeting; handover boundaries*
Operational Readiness (O)	*Are people, equipment and temporary systems ready before exposure begins?*	*Competence/orientation; rescue equipment; scaffold tags; ventilation; calibrated monitors; inspection status*

3.6 Critical-control overrides and numerical claim boundaries

Critical-control failures must be addressed under the site's existing authorization, stop-work and escalation arrangements, regardless of any index. Figure 3 separates that operational obligation from evidence assessment. The author-proposed numerical bands and dimension gate are retained only in Appendix C as uncalibrated examples. They must not define work authorization, acceptable residual risk or a project's safety status.

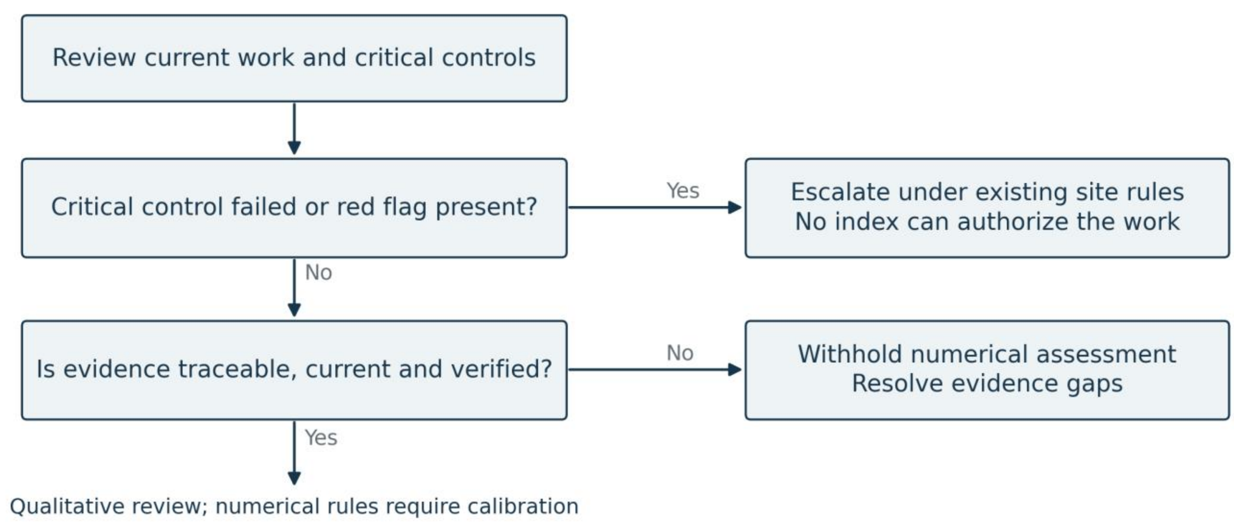

Figure 3. Critical-control review and evidence eligibility. Operational decisions remain governed by site procedures.

3.7 Construct status and reproducibility

No inter-rater reliability, construct validity, criterion validity, predictive validity or causal effect has been demonstrated. The author's operational involvement creates selection and confirmation risks that conservative prose or coding rules cannot eliminate. The retrospective material illustrates a possible evidence structure; it does not establish that a second coder would reach the same judgment.

Appendix B2 specifies a future independent-coding protocol. Two or more independent raters must examine the same frozen, de-identified source package before reconciliation. Agreement statistics, missingness, code distributions and disagreement causes must be reported. Expert calibration and prospective replication are subsequent requirements, not activities completed in this manuscript.

4. Author's Retrospective Illustration of Evidence Mapping

This section reports the author's interpretation of possible evidence sources and the reasons numerical case scoring is withheld. It is not a validation results section. Reported observations, derived exposure calculations and proposed future indicators remain explicitly separate. All numerical RSRI examples appear only in Appendix C.

4.1 Proposed control architecture

Figure 4 organizes technical barriers, management systems and adaptive processes. This is an author-proposed conceptual architecture informed by the reported work context. It does not demonstrate that the barriers were continuously effective or that the six dimensions are empirically distinct.

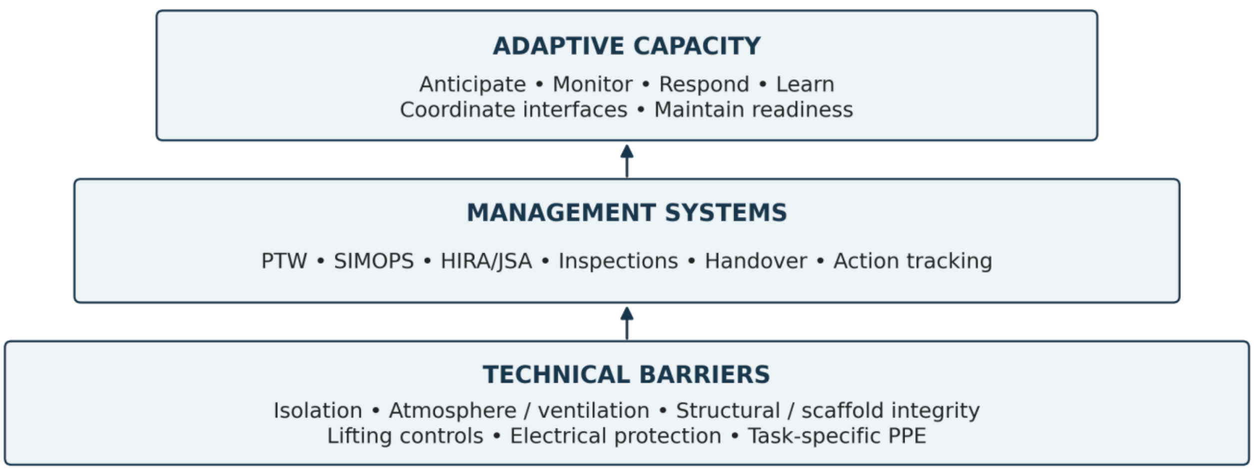

Figure 4. Proposed restoration control architecture; conceptual organization, not a validated causal model.

4.2 Evidence quality and consequences for item coding

Table 7 applies the Table 5 rubric to the available summaries. Partial completeness and timing coexist with no demonstrated independent verification or linked recurrence series. These are limitations of the available evidence, not findings that the underlying work controls failed. The resulting project-level decision is to withhold numerical dimension scores and the RSRI.

The table gives item-specific restrictions rather than retrospectively inventing codes. A ceiling of 2 is conditional: it applies only if a dated, traceable record verifies a defined control but does not demonstrate consistency or effectiveness. Where even that source-level evidence is unavailable, NE remains the appropriate entry. Codes 3–4 cannot be inferred from counts, attendance, staffing or unverified improvement percentages.

Table 7. Author's assessment of summary-evidence quality and item-level scoring restrictions

Evidence family	C	T	V	R	Item-level restriction and missing evidence
Anticipation: look-ahead, SIMOPS review, dynamic HIRA	P	P	ND	ND	A1/A3: NE from summaries. At most 2 if dated source records confirm a defined process; 3–4 require repeated coverage and verified forecast-driven changes.
Monitoring: 24/7 coverage, 567 close-outs, gas/heat logs	P	P	ND	ND	M1/M2/M3: NE for effectiveness. Counts do not justify 3–4; require monitoring denominator, deviations, timely actions and recovery verification.
Response: stop-work, close-outs, drill summaries	P	P	ND	ND	R1/R2: NE for response performance. No standardized timestamps or linked outcomes; at most 2 for a source-verified defined arrangement.
Learning: reported repeat-finding reduction, handovers	P	ND	ND	ND	L1/L2: NE for improvement. The reported reduction cannot support 4 without baseline, follow-up denominator and linked repeat-event history.
Interfaces: 54-permit snapshot, attendance, coordination	P	ND	ND	ND	I1/I2: NE for interaction-control effectiveness. Permit count is context; require conflict identification, action, verification and a common review window.
Readiness: orientation, scaffold, rescue and ventilation summaries	P	ND	ND	ND	O1/O2/O3: NE for coverage. At most 2 for verified requirements; 3–4 require sampled start-of-work checks and documented corrective feedback.

Night-work close-outs could inform M1 (verification), R1 (response) and L1 (recurrence) only if each finding is linked to control status, corrective action, verification and subsequent exposure. Heat logs could inform M3 (threshold action) and R1 only if measurements are linked to a threshold, a timed work/rest or control change, and a recovery check. Counts of close-outs or heat logs establish none of those links by themselves.

No "weakest dimension" or comparative maturity ranking is inferred from these summaries. Appendix C is a standalone arithmetic illustration; it is not the numerical outcome of Table 7 and should not be used as a baseline for this project.

4.3 Descriptive exposure and permit context

Figure 5 uses the dated exposure series and a separate permit snapshot. From 15 to 20 June, reported manpower and daily work-hours increased by 22.5%, while HSE staffing increased by 12%. The reported 20 June manpower-to-HSE-officer ratio was 44.5:1, a descriptive calculation rather than a staffing standard. These figures describe assurance demand and do not enter a case RSRI.

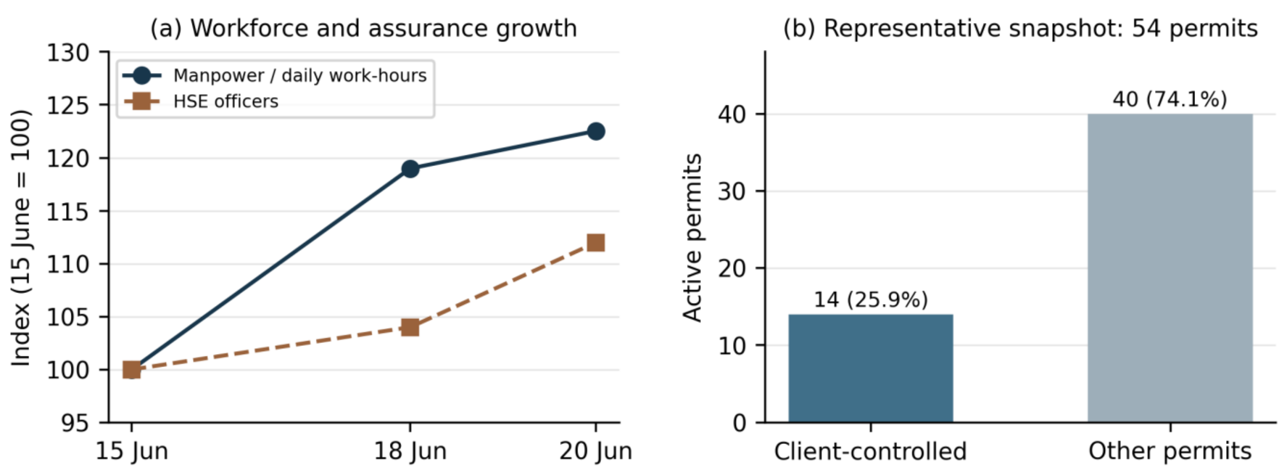

Figure 5. Reported exposure series and separate permit snapshot; context only, not resilience measurements.

4.4 Limits of the reported injury-free period

The reported 1,636,524 cumulative work-hours on 20 June and the later approximately 2.7 million milestone are exposure context. No recordable injury was reported in the supplied summaries, but event definitions, reporting completeness and the exact common observation window have not been independently audited. The absence of reported harm is not used to validate control effectiveness or RSRI.

5. Discussion

5.1 Combining compliance and adaptability

The primary theoretical implication is that Safety-I and Safety-II should not be treated as competing philosophies in high-hazard restoration. Compliance barriers remain non-negotiable: a resilient system cannot compensate for an absent isolation, an untested atmosphere, or an uncertified lifting arrangement through improvisation. At the same time, strict procedure alone is insufficient when the environment changes faster than documents are updated. The practical objective is therefore "bounded adaptability": frontline and supervisory teams can adjust sequencing, resourcing, segregation, ventilation, work-rest cycles, monitoring intensity, or pause criteria, but always within defined critical-control boundaries.

This concept is consistent with the distinction between centralized control and guided adaptability described in resilience literature (Provan et al., 2020). The owner organization defines critical constraints and assurance requirements, while the work system retains enough flexibility to respond to real conditions. The RSRI helps management identify whether the enabling capabilities for this adaptation - monitoring, communication, learning, interface coordination and readiness - are healthy.

5.2 Permit Interaction Load as a proposed future metric

PIL was not operationalized in the case: no georeferenced, timestamped permit-pair dataset, observed interaction series or comparison with discovered conflicts was available. The case's 54-permit snapshot illustrates the scale of a possible search space only: $C = N(N - 1)/2$ gives 1,431 unordered pairs. These are potential pairs, not observed conflicts. Figure 6 describes a proposed screening workflow.

The proposed PIL is Σw_ij for $i < j$, with author-defined interaction weights 0–3. A weight of 0 denotes no material interaction; 1 denotes local coordination without a shared critical barrier or credible escalation pathway; 2 denotes an interaction requiring a specific coordinating control; and 3 denotes potential degradation of another activity's critical control or a high-consequence escalation pathway. These categories require independent calibration before use as a risk-ranking rule.

For $K > 0$, a proposed normalized measure is $nPIL = PIL/(3K)$, where K counts materially interacting screened pairs. It ranges from 0 to 1; if $K = 0$, nPIL is undefined and should be reported as not applicable, alongside $PIL = 0$. Neither measure is an incident probability. Because normalization may conceal the number of interactions, N, K, category counts and the interaction descriptions must accompany any future value.

A future implementation would screen each unordered pair across time, spatial zone, shared isolation/energy, access/egress, lifting/vehicle envelope, atmospheric influence and emergency dependency. Lack of immediate spatial overlap must not automatically exclude a shared-system or atmospheric interaction. The coordinating function would review weights after scope or plant-state changes and remove inactive permits. Validation would compare independently coded interactions with subsequently discovered conflicts and verified control changes.

The worked branch in Figure 6 is hypothetical: selecting 24 pairs with 12 low, 8 medium and 4 high interactions gives $PIL = 40$ and $nPIL = 0.556$. No such 24-pair subset was observed or coded in this project. The example demonstrates a proposed calculation only.

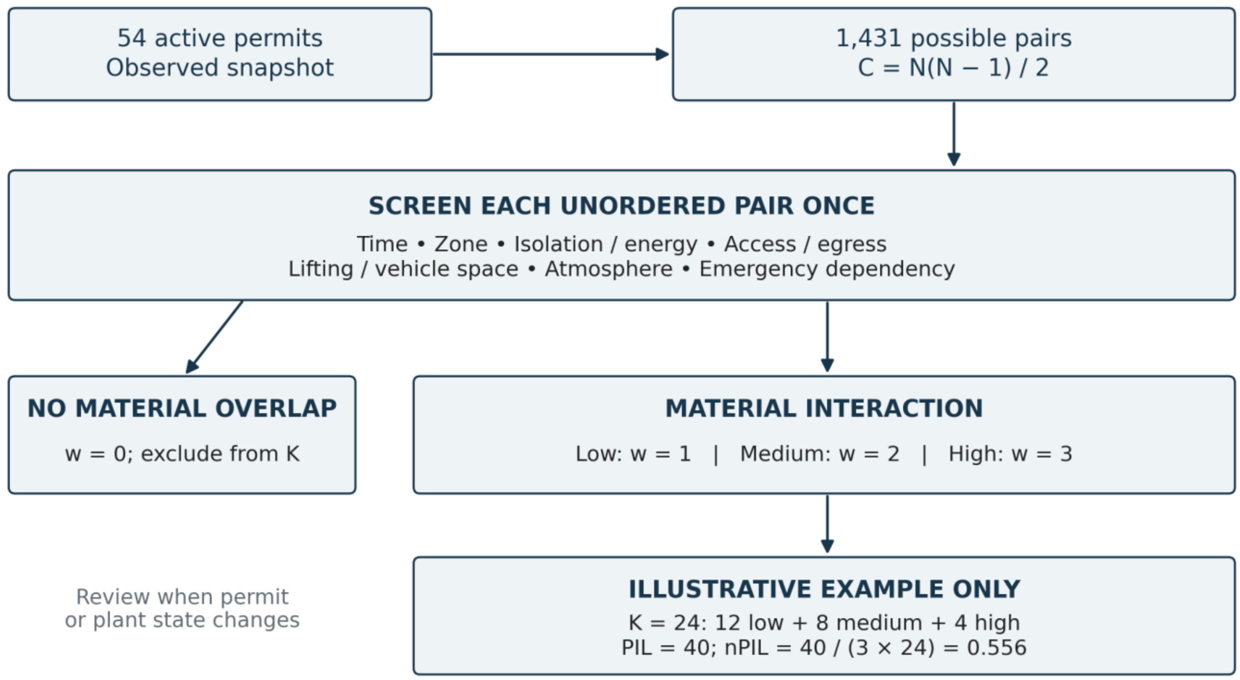

Figure 6. Proposed future PIL workflow. All pair categories and the 24-pair branch are hypothetical and unvalidated.

5.3 What the numerical examples can and cannot show

Appendix C shows that author-assigned dimension inputs can be aggregated and subjected to an arithmetic weight-perturbation exercise. This checks calculation behavior only. It does not justify the weights, establish interval-scale measurement, resolve missing evidence, or demonstrate that the proposed dimensions measure distinct capabilities.

No expert ranking, Delphi round or AHP exercise has been completed. Future elicitation should predefine expert eligibility, elicitation questions, disagreement handling and the criteria for retaining or revising dimensions. Thresholds require separate evaluation of decision consequences and independent control-performance evidence; they cannot be justified by a stable composite alone.

5.4 Operational implications within current site controls

Safety-I controls and Safety-II concepts can inform a qualitative review of changing work. Isolation, gas testing, lifting controls and rescue readiness remain governed by existing approved arrangements. Weak signals, heat-stress adaptations and shift handovers become evidence of adaptive capability only when linked to documented action and verification. RSRI does not create an alternative authorization route.

At this stage, the proposed framework may be used to structure questions, identify missing evidence and plan a feasibility study. It must not be represented as a validated assurance dashboard, used to rank contractors numerically or cited as proof that a project is resilient.

5.5 Proposed 90-day feasibility pilot

Figure 7 outlines a proposed process-feasibility pilot, not a tested implementation programme. Existing PTW, SIMOPS, handover and readiness records would supply a frozen evidence package and a log of practical coding difficulties. Field verification and independent assessment would determine whether defensible item scores are possible. Completion of 90 days would not itself authorize broader numerical deployment.

Figure 7. Proposed 90-day feasibility pilot; completion does not establish validation or authorize score-based assurance.

The initial review display should show source coverage, unresolved critical-control issues and missing evidence. Dimension scores would remain withheld until coding reproducibility, coverage rules and weighting/threshold choices have been examined. A research pilot must not replace existing PTW decisions or operational supervision.

5.6 Wording feedback and required validation

5.6.1 Nine-practitioner wording and face-level check

Nine HSE practitioners reviewed the 30 proposed questionnaire statements and could leave comments. This was a small convenience exercise, not a structured content-validity panel or evidence-coding study. The response anchors described evidence maturity (0–4), rather than dedicated clarity or relevance ratings. Accordingly, the numerical responses are not interpreted as measures of wording quality, practical relevance, project capability or the framework's factor structure.

Comments suggested explanations for low ratings, clearer supervisor/HSE understanding of corrective and preventive action, stronger HAZOP awareness and improved digital documentation. These observations inform wording and pilot planning only; they do not validate RSRI. The full proposed instrument remains in Appendix A.

5.6.2 Validation pathway

Figure 8 separates reproducibility, expert calibration and subsequent empirical testing. For ordinal item codes, a prospective study should pre-specify weighted Cohen's κ for two raters or ordinal Krippendorff's α when appropriate, alongside raw agreement, uncertainty intervals, missingness and code distributions (Cohen, 1960; Krippendorff, 2018). No agreement statistic is calculated here. Appendix B2 specifies the evidence-package and independent-coding requirements.

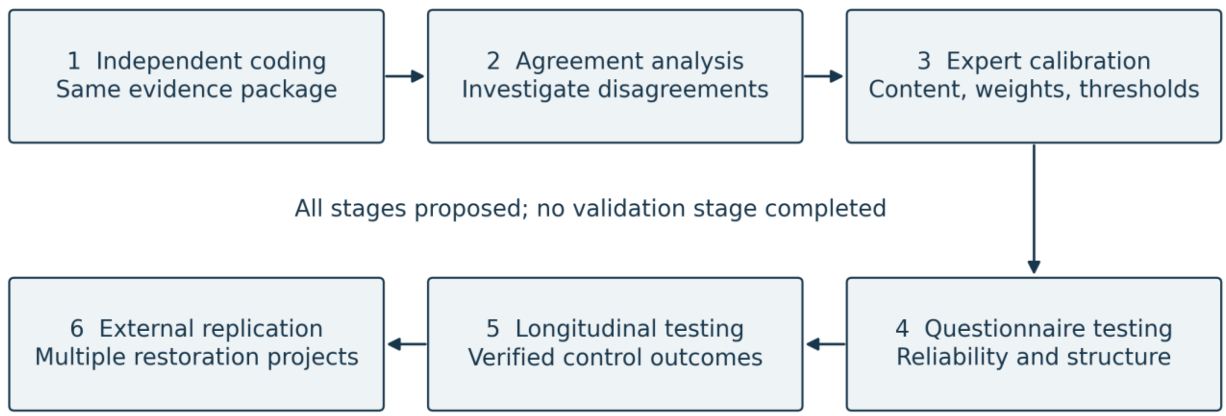

Figure 8. Required future validation sequence; none of these stages is claimed as completed by this study.

5.7 Limitations

The principal limitations are single-author selection and interpretation of operational summaries, no independently coded source package, temporally heterogeneous records, incomplete denominators and recurrence data, and no demonstrated reproducibility. The dimension structure, evidence-quality rubric, coverage policy, weights, bands and PIL categories are author-proposed. The nine-practitioner exercise supplies wording comments only, and its evidence-maturity anchors do not measure clarity or relevance. No case RSRI, predictive validity, causal improvement or generalizable maturity classification is established. Documentation volume can create false assurance, so any future study must test source quality against independent field-control evidence.

The next step is a bounded prospective feasibility study with a common review window, source-level traceability, independent coding and a pre-specified analysis plan. Only then can the framework's measurement properties and practical costs be assessed.

6. Conclusion

This paper proposes a restoration-focused organization of resilience-related evidence across four core potentials and two enabling capabilities. The retrospective illustration identifies potential data sources and demonstrates why incomplete provenance, mixed reporting periods and absent independent verification prevent defensible numerical project scoring. The contribution is the framework and its explicit evidence constraints, not a measured resilience result.

Numerical RSRI values are confined to an author-assigned worked example in Appendix C. PIL remains a proposed future metric, and practitioner comments support wording refinement only. Independent double-coding, expert calibration of content and decision rules, and prospective multi-project testing are required before RSRI can support measurement claims or score-based assurance.

Statements and Declarations

Data Availability Statement

The case material comprises anonymized summaries of internal project HSE records. Raw source records are not publicly distributed because of commercial and operational sensitivity. No standardized source-record package or independently verified item-code dataset is released with this paper. Practitioner comments are summarized anonymously; the response counts are not treated as validation evidence. Any future sharing of de-identified records requires organizational approval.

Ethics Statement

The study uses de-identified operational summaries. Nine HSE practitioners voluntarily supplied questionnaire feedback; names and sensitive personal information are not reported. No independent case-coding exercise or intervention study was conducted.

Conflicts of Interest

The author declares no conflict of interest in relation to the conceptual framework presented in this manuscript. Any employer or client views should not be inferred from the paper unless formally approved.

Acknowledgments

The author acknowledges the contributions of project HSE, operations, construction, permit, and contractor teams whose routine safety practices and records informed the applied case analysis. Organization names are withheld for confidentiality.

Funding

Funding: This research received no external funding.

Author Contributions (CRediT)

Conceptualization, methodology, investigation, data curation, formal analysis, visualization, writing - original draft, and writing - review and editing: Shajahan Abdul Salam.

Confidentiality Statement

All organization, client, contractor, and precise location identifiers are anonymized. Proprietary drawings, client-specific technical details, and directly identifying project records are not reproduced in the manuscript.

ORCID iD: https://orcid.org/0009-0008-1568-0041

Publisher's Note: All claims expressed in this article are solely those of the author and do not necessarily represent those of the affiliated organization, or those of the publisher, the editors and the reviewers.

References

[1]. Adekunle, S., Onatayo, D., Ejohwomu, I., & Aigbavboa, C. (2024). Examining challenges in adopting safety leading indicators for construction projects in South Africa. Frontiers in Built Environment, 10, 1441941. https://doi.org/10.3389/fbuil.2024.1441941
[2]. Alruqi, W. M., Hallowell, M. R., & Techera, U. (2018). Safety climate dimensions and their relationship to construction safety performance: A meta-analytic review. Safety Science, 109, 165–173. https://doi.org/10.1016/j.ssci.2018.05.019
[3]. American Petroleum Institute. (n.d.). API Recommended Practice 754, 3rd edition. https://www.api.org/products-and-services/standards/important-standards-announcements/754
[4]. Arifin, K., Ahmad, M. A., Abas, A., & Mansor Ali, M. X. (2023). Systematic literature review: Characteristics of confined space hazards in the construction sector. Results in Engineering, 18, 101188. https://doi.org/10.1016/j.rineng.2023.101188
[5]. Booth, M., & Butler, J. D. (1992). A new approach to permit to work systems offshore. Safety Science, 15(4–6), 309–326. https://doi.org/10.1016/0925-7535(92)90022-R
[6]. Center for Chemical Process Safety. (2022). Process safety metrics guide for leading and lagging indicators (Version 4.1). AIChE.
[7]. Chen, Y., McCabe, B., & Hyatt, D. (2018). A resilience safety climate model predicting construction safety performance. Safety Science, 109, 434–445. https://doi.org/10.1016/j.ssci.2018.07.003
[8]. Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

[9]. Eustáquio, C. C. L., Cruz, F. M., Araújo, R. P., & Santos, A. C. C. (2024). Confined spaces at construction sites. International Journal of Environmental Engineering, 12(3), 271–286. https://doi.org/10.1504/IJEE.2024.137357

[10]. Fan, H., Enshaei, H., & Jayasinghe, S. G. (2021). Safety philosophy and risk analysis methodology for LNG bunkering simultaneous operations (SIMOPs): A literature review. Safety Science, 136, 105150. https://doi.org/10.1016/j.ssci.2020.105150

[11]. Fan, H., Enshaei, H., & Jayasinghe, S. G. (2023). Dynamic quantitative risk assessment of LNG bunkering SIMOPs based on Bayesian network. Journal of Ocean Engineering and Science, 8(5), 508–526. https://doi.org/10.1016/j.joes.2022.03.004

[12]. Golabchi, H., Abellanosa, A. D., Lefsrud, L., Pereira, E., & Mohamed, Y. (2024). A comprehensive systematic review of safety leading indicators in construction. Safety Science, 172, 106433. https://doi.org/10.1016/j.ssci.2024.106433

[13]. Guo, B. H. W., & Yiu, T. W. (2016). Developing leading indicators to monitor the safety conditions of construction projects. Journal of Management in Engineering, 32(1), 04015016. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000376

[14]. Health and Safety Executive. (2005). Guidance on permit-to-work systems: A guide for the petroleum, chemical and allied industries (HSG250). https://www.hse.gov.uk/pubns/priced/hsg250.pdf

[15]. Hinze, J., Thurman, S., & Wehle, A. (2013). Leading indicators of construction safety performance. Safety Science, 51(1), 23–28. https://doi.org/10.1016/j.ssci.2012.05.016

[16]. Hollnagel, E. (2014). Safety-I and Safety-II: The past and future of safety management. Ashgate/Routledge.

[17]. Hollnagel, E. (2017). Safety-II in practice: Developing the resilience potentials. Routledge.

[18]. Hollnagel, E., Pariès, J., Woods, D. D., & Wreathall, J. (Eds.). (2011). Resilience engineering in practice: A guidebook. Ashgate.

[19]. Kim, Y., & Jung, S. (2011). A practical framework for mandatory job safety analysis embedded in the permit-to-work system and application to gas industry. Journal of Chemical Engineering of Japan, 44(12), 976–988. https://doi.org/10.1252/jcej.11we156

[20]. Krippendorff, K. (2018). Content analysis: An introduction to its methodology (4th ed.). SAGE Publications.

[21]. Marucco, D. (2016). Simultaneous operations risk assessment. Chemical Engineering Transactions, 53, 115–120. https://doi.org/10.3303/CET1653020

[22]. Mollo, L. G. (2026). Resilient health and safety performance on construction sites from a Safety-II perspective. Discover Civil Engineering, 3, 109. https://doi.org/10.1007/s44290-026-00509-8

[23]. Peñaloza, G. A., Formoso, C. T., & Saurin, T. A. (2021). A resilience engineering-based framework for assessing safety performance measurement systems: A study in the construction industry. Safety Science, 142, 105364. https://doi.org/10.1016/j.ssci.2021.105364

[24]. Pika, A., Ter, A., Perrons, R. K., & Grossmann, G. (2021). Using big data to improve safety performance: An application of process mining to enhance data visualisation. Big Data Research, 25, 100210. https://doi.org/10.1016/j.bdr.2021.100210

[25]. Provan, D. J., Woods, D. D., Dekker, S. W. A., & Rae, A. J. (2020). Safety II professionals: How resilience engineering can transform safety practice. Reliability Engineering & System Safety, 195, 106740. https://doi.org/10.1016/j.ress.2019.106740

[26]. Shaikh, A. Y., Osei-Kyei, R., & Hardie, M. (2020). A critical analysis of safety performance indicators in construction. International Journal of Building Pathology and Adaptation, 39(3), 547–580. https://doi.org/10.1108/IJBPA-03-2020-0018

[27]. Xu, J., Cheung, C., Manu, P., & Ejohwomu, O. (2021). Safety leading indicators in construction: A systematic review. Safety Science, 139, 105250. https://doi.org/10.1016/j.ssci.2021.105250

[28]. Xu, J., Cheung, C., Manu, P., Ejohwomu, O., & Too, J. (2023). Implementing safety leading indicators in construction: Toward a proactive approach to safety management. Safety Science, 157, 105929. https://doi.org/10.1016/j.ssci.2022.105929

[29]. Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE Publications.

- **Appendix A. Proposed Restoration Safety Resilience Questionnaire**

Proposed response scale: 0 = No evidence, 1 = Initial / weak evidence, 2 = Developing / partial evidence, 3 = Established / consistent evidence, 4 = Strong / mature evidence. These are evidence-maturity anchors, not clarity or relevance ratings. The nine-practitioner exercise provided wording feedback only and did not validate this scale. Future testing should add a separate "not evaluable" option and distinguish missing evidence from demonstrated absence.

No.	***Dimension***	***Statement***	***Rating***
1	*Anticipation*	*High-risk work for the next three days is reviewed for interactions before permits are issued.*	*0 1 2 3 4*
2	*Anticipation*	*Changes in plant condition, isolations, access routes and temporary systems are considered before work starts.*	*0 1 2 3 4*
3	*Anticipation*	*SIMOPS conflicts are normally identified during planning rather than after work begins.*	*0 1 2 3 4*

No.	***Dimension***	***Statement***	***Rating***
4	*Anticipation*	*Weather, heat and other environmental conditions are considered when work sequencing is planned.*	*0 1 2 3 4*
5	*Anticipation*	*Frontline teams are encouraged to raise credible "what if" scenarios before starting work.*	*0 1 2 3 4*
6	*Monitoring*	*Critical controls are verified in the field rather than assumed from paperwork.*	*0 1 2 3 4*
7	*Monitoring*	*Atmospheric conditions are monitored at the frequency required by the actual risk.*	*0 1 2 3 4*
8	*Monitoring*	*Night-shift and low-supervision periods receive adequate safety monitoring.*	*0 1 2 3 4*
9	*Monitoring*	*Changes to temporary ventilation, scaffolds, access or exclusion zones are detected promptly.*	*0 1 2 3 4*
10	*Monitoring*	*Weak signals and near misses are captured even when no injury occurs.*	*0 1 2 3 4*
11	*Response*	*Workers can stop work when conditions differ from the permit or risk assessment.*	*0 1 2 3 4*
12	*Response*	*Critical safety issues are escalated quickly to people with authority to act.*	*0 1 2 3 4*
13	*Response*	*Control measures are adjusted promptly when monitoring shows deterioration.*	*0 1 2 3 4*
14	*Response*	*Emergency and rescue arrangements are practical for the actual work location.*	*0 1 2 3 4*
15	*Response*	*Corrective actions prioritize potential severity, not only administrative due dates.*	*0 1 2 3 4*
16	*Learning*	*Observation and near-miss findings are analyzed for recurrence and common causes.*	*0 1 2 3 4*
17	*Learning*	*Lessons from one contractor or work area are shared with other affected teams.*	*0 1 2 3 4*
18	*Learning*	*Shift handovers transfer unresolved hazards and temporary controls effectively.*	*0 1 2 3 4*
19	*Learning*	*Risk assessments and work methods are revised when field experience shows a gap.*	*0 1 2 3 4*
20	*Learning*	*Management reviews whether corrective actions prevented the issue from recurring.*	*0 1 2 3 4*
21	*Interface Coordination*	*Permit issuers understand the hazardous interactions among simultaneous jobs.*	*0 1 2 3 4*
22	*Interface Coordination*	*Client and contractor teams share a common view of current high-risk work.*	*0 1 2 3 4*
23	*Interface Coordination*	*Exclusion zones, lifting routes, access routes and emergency routes are coordinated across work groups.*	*0 1 2 3 4*
24	*Interface Coordination*	*Conflicting work is resequenced when segregation or additional controls are not sufficient.*	*0 1 2 3 4*
25	*Interface Coordination*	*Permit density and work congestion are treated as risk factors in themselves.*	*0 1 2 3 4*
26	*Operational Readiness*	*Workers are competent and oriented for the task and site hazards before exposure begins.*	*0 1 2 3 4*
27	*Operational Readiness*	*Critical temporary systems such as ventilation, lighting and access are verified before use.*	*0 1 2 3 4*
28	*Operational Readiness*	*Rescue equipment and trained personnel are available before confined-space or work-at-height tasks begin.*	*0 1 2 3 4*
29	*Operational Readiness*	*Gas detectors, lifting gear and other safety-critical equipment are within inspection/calibration status.*	*0 1 2 3 4*

No.	Dimension	Statement	Rating
30	Operational Readiness	High-risk work does not start when required supervision or specialist support is unavailable.	0 1 2 3 4

- **Appendix B. Proposed RSRI Evidence-Coding Checklist**

Code	Evidence test	Typical source
A1	3-day look-ahead exists and covers high-risk activities / SIMOPS	Meeting record / plan
A2	Plant-state, isolation and access changes are forecast	Permit / isolation / coordination record
A3	Dynamic HIRA/JSA is revised when conditions change	Revision history / field evidence
M1	Critical controls are field-verified	Inspection / audit record
M2	Atmospheric monitoring is appropriate to risk	Gas-test logs / continuous monitor record
M3	Heat / environmental thresholds trigger action	Monitoring logs / work-rest changes
R1	Stop-work and escalation are used when required	Stop-work / incident / observation record
R2	Emergency and rescue performance is tested	Drill record / response time
L1	Repeat observations are tracked	Trend report
L2	Lessons are transferred across shifts / contractors	Handover / briefing evidence
I1	Active permit interactions are reviewed	SIMOPS matrix / permit board
I2	High-interaction work is segregated or resequenced	Coordination record
O1	Competence and orientation are current	Training matrix
O2	Temporary safety-critical systems are inspected and ready	Scaffold / ventilation / electrical / access inspection
O3	Rescue and specialist support are available before start	Readiness checklist

- **Appendix B1. RSRI 0–4 Coding Decision Rules**

For a future standardized assessment, raters must cite the source record, applicable period and reason for each item code. NE denotes insufficient evidence and is separate from 0, which requires demonstrated absence or contradiction. A procedure alone cannot establish implementation. The lower defensible code applies where evidence is conflicting; codes 3–4 require coverage and verification beyond activity volume. Table B1 sets out the proposed anchors; Table 7 explains why the current summaries cannot establish high-maturity item codes.

Table B1. Decision rules for RSRI evidence coding

Code	Minimum evidence standard	Typical decision rule
0 – Absent / contradictory	No control evidence, or records show the required control was bypassed/failed without recovery	Use 0 only when absence/contradiction is demonstrable
1 – Ad hoc	Isolated action or person-dependent practice; no stable process or repeatable record	One-off good practice does not justify routine maturity
2 – Defined / partially evidenced	Procedure or recurring activity exists, but consistency, coverage, or effectiveness is unclear	Default ceiling when only procedure/activity evidence is available

Code	*Minimum evidence standard*	*Typical decision rule*
3 – Routinely implemented	*Repeated evidence across the review period with defined ownership and field verification*	*Requires more than one evidence point or a demonstrably complete record set*
4 – Implemented with feedback/adaptation	*Routine implementation plus trend review, feedback, learning, or documented control adjustment*	*Requires evidence that information changed subsequent planning/control*

The checklist above is a candidate coding aid, not a completed rating sheet. No item-level case-code dataset is claimed.

- **Appendix B2. Proposed independent-coding protocol**

This protocol has not been executed. Freeze a de-identified package for one review window, with source identifiers, permit/shift coverage, an expected-control inventory, contradictions and a missing-data log. Exclude the assigned Appendix C scores.

At least two independent coders with relevant expertise would receive the same codebook and separate training examples. Before discussion, each would record applicability, source IDs, quality labels, NE status, codes and reasons. Preserve original ratings separately from subsequent consensus.

Pre-specify ordinal agreement: weighted Cohen's κ for two raters or ordinal Krippendorff's α where appropriate (Cohen, 1960; Krippendorff, 2018). Report the rating matrix, raw agreement, category prevalence, uncertainty and disagreement causes. Report missingness and applicability agreement separately. Revise the codebook and retest on a fresh package before making reproducibility claims.

- **Appendix C. Worked example only under author-proposed assumptions**

ILLUSTRATIVE ONLY — SINGLE-CODER. This appendix is an arithmetic teaching example, not a result of case evidence scoring. The six dimension inputs were assigned by the author in the earlier illustration. No recoverable item-code matrix establishes them, and they are not the output of the evidence-quality assessment in Table 7. They must not be interpreted as the project's performance, maturity or probability of avoiding incidents.

- **C1. Input profile and aggregation**

Table C1. Author-assigned numerical inputs and uncalibrated example weights

Dimension	**Illustrative input /100**	**Author-proposed weight**	**Weighted contribution**
Anticipation	85	0.20	17.00
Monitoring	90	0.20	18.00
Response	82	0.15	12.30
Learning	75	0.15	11.25
Interface Coordination	88	0.20	17.60
Operational Readiness	85	0.10	8.50

Using only these assigned inputs, RSRI = 0.20(85) + 0.20(90) + 0.15(82) + 0.15(75) + 0.20(88) + 0.10(85) = 84.65/100. This precision belongs to the arithmetic and does not imply corresponding measurement accuracy. It is not an empirical project score.

In a future scored dataset, each dimension would instead be calculated as (mean of admissible 0–4 item codes / 4) × 100. For example, purely hypothetical codes [3, 2, 3, 2] yield a mean of 2.5 and a dimension value of 62.5/100. These four invented teaching codes are not observations and do not reconstruct the six assigned inputs in Table C1.

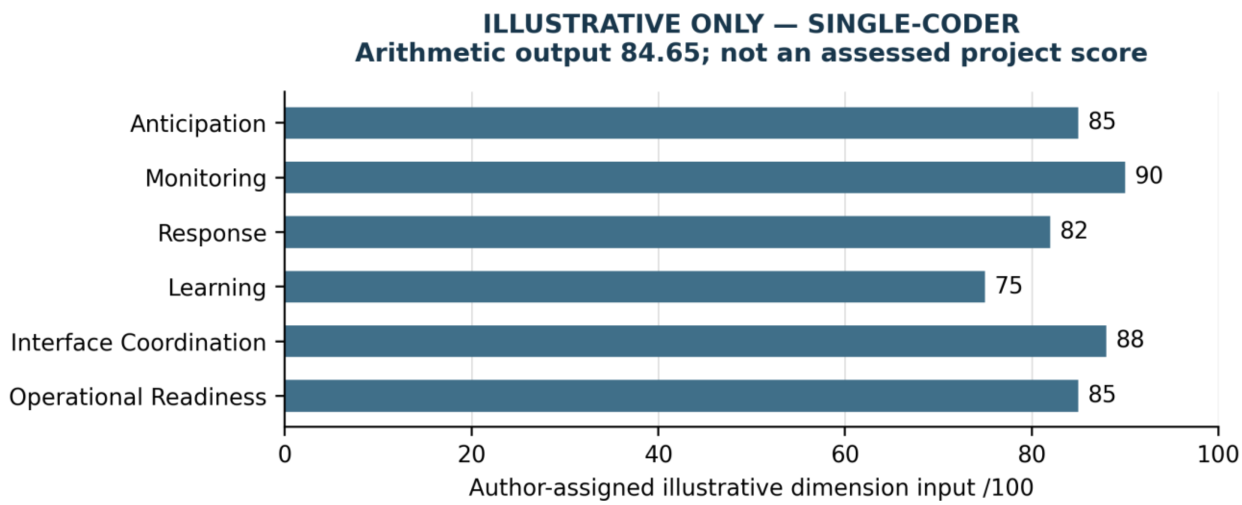

Figure C1. Illustrative only — single-coder author-assigned profile; not an assessed project result.

- **C2. Arithmetic sensitivity only**

Equal weighting of the assigned profile gives 84.17/100. Varying each weight individually by ±25%, with the other weights proportionally rescaled to maintain a total of 1, gives 84.22–85.08/100. The widest range follows changes to Learning, the lowest assigned input. This exercise establishes only arithmetic behavior for these assumptions, not reliability, content validity or robustness to erroneous item scores.

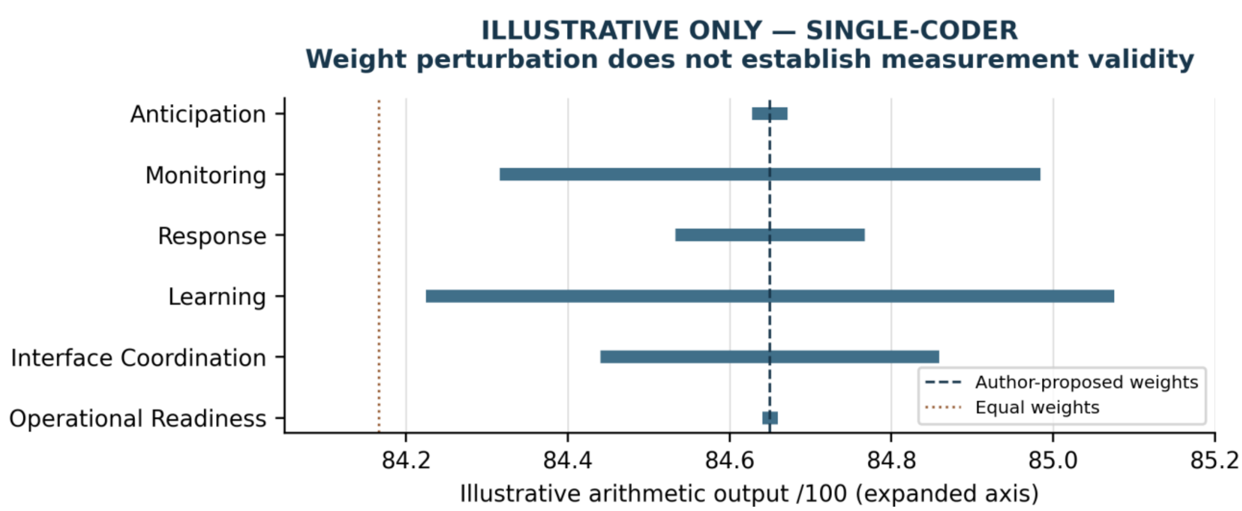

Figure C2. Illustrative only — single-coder weight perturbation with fixed author-assigned inputs.

- **C3. Uncalibrated bands retained for methodological transparency**

Table C2 preserves the original author-proposed labels solely to expose design choices requiring testing. The intervals below are made non-overlapping for continuous scores. Their boundaries and labels have no validated empirical meaning; even a high arithmetic output cannot establish safe conditions.

Table C2. Illustrative classification assumptions only; not operational decision thresholds

Hypothetical score interval	Author-proposed label	Status
$0 \le x < 50$	Critical / brittle	Unvalidated example
$50 \le x < 65$	Reactive	Unvalidated example
$65 \le x < 80$	Managed	Unvalidated example
$80 \le x < 90$	Resilient	Unvalidated example
$90 \le x \le 100$	High resilience / continuously learning	Unvalidated example

The earlier example gate of any critical-control dimension below 60 was also author-proposed and is not calibrated or applied to the case. Critical failures require action under existing site rules without waiting for a score. Future validation must define which item/control conditions trigger review, how evidence gaps are handled and the consequences of false reassurance. No maturity label is assigned to the case in the main paper.