Information & Computer Security

Understanding incident response readiness through human and system telemetry

Muntathar Abid and Priyadarsi Nanda
Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia, and

Manoranjan Mohanty
Dietrich College of Humanities and Social Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Received 9 February 2026
Revised 28 April 2026
Accepted 17 May 2026

Abstract

Purpose – Small-to-medium enterprises (SMEs) remain vulnerable to cyber incidents because of resource constraints. While incident response plans exist, they are often untested, creating a critical readiness gap. This paper aims to introduce the Incident Response Readiness Score (IRRS), a scenario-based framework designed to empirically evaluate organisational incident response (IR) capability under simulated conditions.

Design/methodology/approach – The IRRS applies a structured scoring rubric calibrated through a Scenario Risk Index to evaluate performance. The authors further introduce human-centric telemetry – decision latency, communication entropy and authority drift – which operate as diagnostic lenses to explain observed readiness outcomes and identify organisational friction without altering IRRS scoring.

Findings – The study diagnosed an SME with a "Reactive" maturity level, revealing a stark contrast between technical potential and operational reality. In spite of modern tooling, response capability was compromised by human–system friction. Telemetry isolated extended decision latency and significant authority drift as root causes. These findings confirm that readiness is defined by decision clarity rather than tool ownership and that the IRRS successfully exposes cognitive failures that static audits miss.

Originality/value – This paper presents the first empirical framework specifically designed to quantify SME IR readiness using risk-weighted simulations. Unlike static compliance audits, IRRS isolates the gap between documented policy and operational execution. The introduction of diagnostic telemetry offers a novel method for distinguishing between technical deficits and human-process failures.

Keywords Cybersecurity, Incident response evaluation, Incident response readiness scoring, Incident response planning, Human-centric telemetry, Scenario-based simulation

Paper type Research paper

1. Introduction

Cyber incidents now pose an existential threat for small-to-medium enterprises (SMEs). In spite of regulatory frameworks in jurisdictions such as Australia, the UK and the European Union mandating breach notification and due diligence obligations regardless of organisational size, SMEs disproportionately struggle to implement robust cybersecurity controls because of inherent resource constraints (Verizon, 2025). With limited access to

Information & Computer Security
Emerald Publishing Limited
2056-4961
DOI 10.1108/ICS-02-2026-0068

dedicated security personnel, enterprise-grade tooling or internal cyber governance functions, SMEs are often forced to prioritise baseline compliance over true operational preparedness. As a result, incident response plans (IRPs) are frequently developed as regulatory artefacts and remain untested, usually archived until a real breach occurs only to find they fail at the first step (Goings *et al.*, 2016). This compels urgent improvisation under uncertainty conditions, pressure and competing business priorities. This gap has been repeatedly cited as a critical vulnerability within the SME sector, particularly given the high prevalence of successful attacks targeting this segment (Australian Cyber Security Centre (ACSC), 2021). Major cyber incidents have led to financial collapse for over 60% of European SMEs, with a majority closing within half a year of the breach (Boswell, 2022). Although traditional readiness assessment approaches such as ISO/IEC 27035 audits, policy compliance reviews and tabletop exercises are widespread, current incident response (IR) learning processes often fail to integrate people, processes and technological insights dynamically, instead focusing on limited objectives such as service restoration or formal reporting without ensuring broader organisational security learning (Ahmad *et al.*, 2015).

Complicating the matter further, leading cybersecurity maturity models such as the National Institute of Standards and Technology (NIST) Cybersecurity Framework and security incident management maturity model (SIM3) assume availability of substantial security infrastructure, such as security operations centres (SOCs), security information and event management (SIEM) platforms and continuous monitoring regimes-resources often beyond the financial or operational capacity of SMEs. While lightweight alternatives such as IASME Cyber Assurance have emerged (Andy Dodd, 2024), the broader maturity model landscape remains insufficiently aligned with the live operational realities of SMEs (Chidukwani *et al.*, 2024). To address these limitations, this paper introduces the Incident Response Readiness Score (IRRS), a scenario-driven, risk-adjusted framework designed to translate qualitative assessments of IR capability into measurable, repeatable metrics.

The IRRS operationalises real-world scenarios, including ransomware, insider threats and cloud misconfigurations, within live or sandboxed environments. It evaluates IR execution across five domains: procedural fidelity, operational execution, infrastructure integration, team coordination and post-incident learning. Validation and performance of the proposed scheme are weighted using a Scenario Risk Index (SRI) calibrated to both threat complexity and business impact, enabling longitudinal tracking through repeated simulation cycles across a four-tier maturity model ranging from *ad hoc* to adaptive. The framework delivers four key contributions: it enables cost-effective resilience assessment for SMEs; aligns measurement with observed operational behaviour rather than compliance artefacts; demonstrates correlation between readiness scores and organisational characteristics such as training frequency; and structures iterative post-incident learning. Collectively, IRRS provides a defensible strategy for bridging the gap between documented preparedness and operational readiness.

While IRRS quantifies IR performance, readiness scores alone do not explain why response failures or delays arise under operational pressure. Empirical observations from simulated incidents indicate that organisations with comparable technical capabilities may nevertheless exhibit distinct behavioural dynamics driven by cognitive load, coordination friction and ambiguous decision authority. To address these explanatory gaps, the framework is extended with human-centric telemetry metrics that function purely as diagnostic lenses, capturing decision-making and coordination behaviours to contextualise results without altering core IRRS scoring.

2. Literature review

The field of IR in cybersecurity has undergone significant evolution and improvements in recent years, transitioning from *ad hoc*, reactive procedures towards structured, lifecycle-based

approaches guided by authoritative frameworks. Notably, NIST Special Publication 800-61 Revision 3 (National Institute of Standards and Technology (NIST), 2024), ISO/IEC 27035-2:2023 (International Organization for Standardization, 2023) and the SANS Institute's Incident Handler's Handbook (Kral, 2012) provide comprehensive guidance across critical incident management phases, encompassing preparation, detection, containment, eradication and recovery. However, in spite of their widespread recognition, small and medium-sized enterprises (SMEs) commonly face significant barriers to fully operationalising these guidelines because of inherent resource constraints, limited cybersecurity expertise and informal governance structures (Banham, 2017).

To address these limitations, recent scholarship and industry research have turned towards incident response maturity models (IRMMs), designed to systematically evaluate organisational readiness by benchmarking procedural and technical capabilities. Models such as the SIM3 (Open CSIRT Foundation, 2023) and the European Union Agency for Cybersecurity (ENISA) CSIRT Maturity Framework (European Union Agency for Cybersecurity (ENISA), 2022) provide valuable frameworks; however, these tools primarily cater to larger enterprises or national-level teams (e.g. CSIRTs) that possess formalised structures and dedicated response resources. Typically reliant upon qualitative assessments through stakeholder interviews or artefact reviews, these existing maturity models lack empirical rigour, especially in identifying operational bottlenecks and real-world performance discrepancies during live incidents. Such limitations become especially pronounced in SMEs, where incident handling is typically decentralised, *ad hoc* and highly context-dependent, thus reducing the efficacy and relevance of conventional IRMM methodologies in these environments.

In parallel, the adoption of simulation-based training and tabletop exercises has gained traction as practical methods to assess and enhance IR capabilities. Advanced simulation tools, such as MITRE ATT&CK Evaluations (MITRE Engenuity, 2025), Caldera (MITRE Corporation, 2025) and various red/blue team exercises, provide realistic adversary emulation, thereby offering valuable opportunities for organisations to observe and refine their IR processes. Nonetheless, existing simulation tools seldom incorporate structured, quantifiable scoring frameworks to holistically measure IR performance. Furthermore, many of these methods implicitly assume the existence of robust cybersecurity infrastructure, including mature SOCs, integrated SIEM or security orchestration, automation and response platforms and dedicated cybersecurity teams – conditions that rarely exist in SME contexts. Thus, while beneficial in theory, current simulation practices are inadequately tailored to SMEs, limiting their practical utility and leaving a critical gap in accurately measuring and improving IR preparedness within smaller scale organisations. Moreover, although the adoption of risk-adjusted metrics is well established in cybersecurity domains such as vulnerability management, including risk-weighted common vulnerability scoring system (CVSS) methodologies and compliance assurance, including ISO 27001 audits, it remains comparatively unexplored in evaluating IR effectiveness at an organisational process level (National Institute of Standards and Technology (NIST), 2023). Although maturity models informed by capability maturity model integration frameworks have occasionally been applied to IR assessments (Ahmad *et al.*, 2021), their primary focus continues to be the presence and completeness of documented policies, rather than assessing operational fidelity, execution timeliness and integration effectiveness during actual or simulated cyber incidents. Recent research has introduced isolated metrics such as time-to-containment, response success rates and IR drift indicators; however, these approaches remain fragmented and lack integration into a unified, comprehensive and contextually adaptable scoring framework suitable for SMEs. Traditional IR has often been designed with linear, plan-driven process

models that have sequential stages, such as preparation, identification and containment (National Institute of Standards and Technology (NIST), 2024; Kral, 2012). Compounding the issue is that scant attention has been given in the literature to investigating correlations between IR performance metrics and relevant organisational factors, including team size, tooling integration and historical incident experience, all of which significantly affect IR outcomes in resource-constrained SME settings. To address the unique cyber resilience challenges faced by SMEs, tailored frameworks and models have been proposed that emphasise prescriptive detection and IR strategies underpinned by scalable, open-source infrastructure solutions (Ilca *et al.*, 2023). In parallel, scenario-based training frameworks have been developed to confront the socio-technical barriers that inhibit effective IR. These meta-level approaches seek to systematically enhance organisational readiness through structured, context-aware exercises that specifically target deficiencies in communication, coordination and inter-team integration (O'Neill *et al.*, 2021). The integration of artificial intelligence (AI) into IR processes has also been explored, with Oluwawemimo (Oluwawemimo, 2024) analysing the role of AI in enhancing IR capabilities within digital domain SMEs. The use of simulation-based training has been highlighted as an effective method for preparing organisations to handle real-world incidents, with studies emphasising the benefits of immersive and scenario-based exercises (See *et al.*, 2024). While prior work clearly demonstrates the importance of human and organisational factors during security incidents, these insights are typically derived from qualitative analysis or post-incident reflection. As a result, such factors are rarely operationalised within structured readiness measurement frameworks, limiting their use in comparative or simulation-based evaluation. This gap motivates the introduction of lightweight, observable telemetry metrics that can be captured during incident simulations without disrupting response activity.

This review identifies persistent gaps in the evaluation of IR capability within SMEs. First, there is a notable absence of empirically validated approaches for measuring operational execution fidelity. Second, existing methods lack risk-adjusted scoring mechanisms that integrate technical and procedural assessments. Third, the influence of organisational factors – such as staffing levels and toolchain maturity – on response outcomes remains underexplored. Collectively, these limitations constrain the development of actionable assessment frameworks for resource-constrained environments. This study addresses these gaps by introducing a rigorous, quantifiable and context-sensitive approach to IR readiness assessment, empirically validating execution fidelity, linking response maturity to organisational characteristics and establishing a foundation for future work aimed at improving SME IR practices.

3. Baseline scenario and organisational context

To ground the IRRS framework in operational reality, this study uses a composite baseline environment derived from recurring organisational patterns observed across multiple SME security assessments. This baseline does not represent a specific organisation; rather, it synthesises typical operational and governance conditions frequently encountered in SME environments and serves as the reference context for scoring demonstrations, maturity classifications and telemetry analysis presented throughout this paper.

Rather than modelling a single real-world incident, the baseline environment provides a representative organisational context under which IR effectiveness is evaluated. Two representative threat scenarios are used consistently to demonstrate framework applicability across incidents requiring both rapid technical containment and governance-driven decision-making.

3.1 Organisational profile

The baseline organisation represents a composite mid-sized retail enterprise employing approximately 90 staff and operating integrated warehouse logistics alongside e-commerce services. Its technology stack includes Microsoft 365 E5, Microsoft Defender for Endpoint and Cloud App Security, providing foundational detection and endpoint protection capabilities typical of mid-maturity SMEs.

Security monitoring activities are outsourced to a SOC-as-a-Service provider; however, operational containment and business response authority remain internal to the organisation. Governance maturity remains uneven, with an IRP existing but rarely exercised, limited detection capability for lateral movement because of absent correlation rules and insufficient clarity regarding organisational roles during crisis escalation. This configuration reflects organisations that possess reasonable technical tooling but lack the operational readiness necessary to execute IR consistently and efficiently.

3.2 Threat scenarios

Two threat scenarios are used throughout the study to evaluate framework performance across differing operational pressures and decision contexts:

- Phishing-led ransomware (high-intensity, time-critical)

In this scenario, a phishing compromise leads to malware execution and attempted ransomware propagation within the enterprise network. The attack begins when a warehouse floor manager interacts with a malicious email link, resulting in malware execution and subsequent attempts at lateral propagation across internal systems.

The principal operational challenge is rapid isolation of the compromised host before encryption activity impacts shared infrastructure, particularly the central stock control database that underpins warehouse and sales operations. A successful response, therefore, depends on rapid technical execution combined with the authority to disrupt operational services – such as isolating servers or disabling accounts – without procedural delay. This scenario primarily evaluates containment speed and operational decisiveness:

- Insider data exfiltration (low-intensity, high-complexity)

The second scenario models a trusted employee attempting to exfiltrate sensitive pricing data using legitimate system access. In this case, proprietary information is uploaded to personal cloud storage using authorised credentials, thereby bypassing traditional malware-based detection mechanisms.

The operational challenge lies in identifying behaviour that is technically permitted but contextually malicious. Effective response depends on clarity of organisational roles, appropriate escalation pathways and the ability to investigate suspicious behaviour without creating internal governance or human resources conflicts. Consequently, this scenario evaluates judgement, escalation maturity and organisational decision governance:

- The readiness gap

Simulations conducted in environments resembling this composite baseline consistently demonstrate that organisations often struggle to translate technical detection capability into coordinated operational response. While security tooling frequently generates adequate alerts, containment actions are delayed because of unclear escalation pathways, informal decision-making processes and hesitation among technical staff lacking pre-authorised containment authority.

The IRRS framework detailed in the following sections was developed to quantify these operational friction points by transforming qualitative observations of ineffective response into measurable readiness indicators. Rather than merely identifying the occurrence of a failure, the framework isolates the specific behavioural and procedural mechanics that precipitated the outcome. By applying this risk-weighted model to the baseline scenarios described above, the study demonstrates how organisations can systematically diagnose their maturity gaps and progress from reactive improvisation towards measurable and repeatable IR readiness.

4. Incident Response Readiness Score overview and components

This section details the architecture and scoring logic of the IRRS. Designed to measure performance rather than policy compliance, IRRS translates scenario-driven simulated behaviour into a reproducible, risk-aware maturity score for SMEs. By evaluating how effectively organisational resources are applied under stress – across scenarios such as ransomware or insider threats – the framework produces a normalised readiness score (0–100) and a corresponding maturity tier that reflects actual operational IR capability.

The IRRS framework is composed of several interconnected components designed to support realistic simulation, structured evaluation and actionable insight generation. As outlined in Table 1, these include the simulation engine, which initiates sector-relevant incident scenarios; scenario-specific evaluation, which contextualises sub-metric scoring using a weighted risk model; and the IRRS scoring model, which aggregates performance outcomes. The framework also integrates an instrumentation bus to ensure observational fidelity and a maturity curve classification system that interprets results into actionable tiers.

Table 1. IRRS core components

Component	Function
Simulation engine	Executes realistic, threat-informed incident scenarios (table-top or live emulation) that mirror sector-relevant SME threat profiles and are maintained under version-controlled manifests for reproducibility
Scenario-specific evaluation	Assesses participant response to a selected scenario using the Scenario Risk Index (SRI), the scenario weighting guide and a standardised sub-metric scoring scale (0–5). Integrates context-specific risk weighting with evaluator observations to ensure relevance and proportionality
IRRS scoring model	Aggregates scenario scores into a weighted, normalised readiness score (0–100) using predefined sub-metrics across five domains. Designed to enable cross-scenario benchmarking and longitudinal tracking of response capability
Maturity curve classification	Tiered classification that maps IRRS scores into discrete readiness bands and surfaces longitudinal trends across repeated simulations
Instrumentation bus	Passive hooks that record security-tool invocations and human hand-offs during simulations, validating both tooling integration and team-dynamic factors
Feedback and improvement actions	Converts simulation findings into targeted remediation guidance across tooling, procedures and team coordination. Enables iterative readiness uplift and validation of improvement over time

Source(s): Authors' own work

Finally, a feedback and improvement mechanism closes the loop, enabling iterative enhancement of readiness practices.

Figure 1 presents the operational flow of the IRRS framework, illustrating how simulation exercises are translated into measurable readiness insights. The process begins with scenario execution and proceeds through scenario-specific evaluation, risk-adjusted scoring and maturity tier classification. The instrumentation bus enables behavioural data

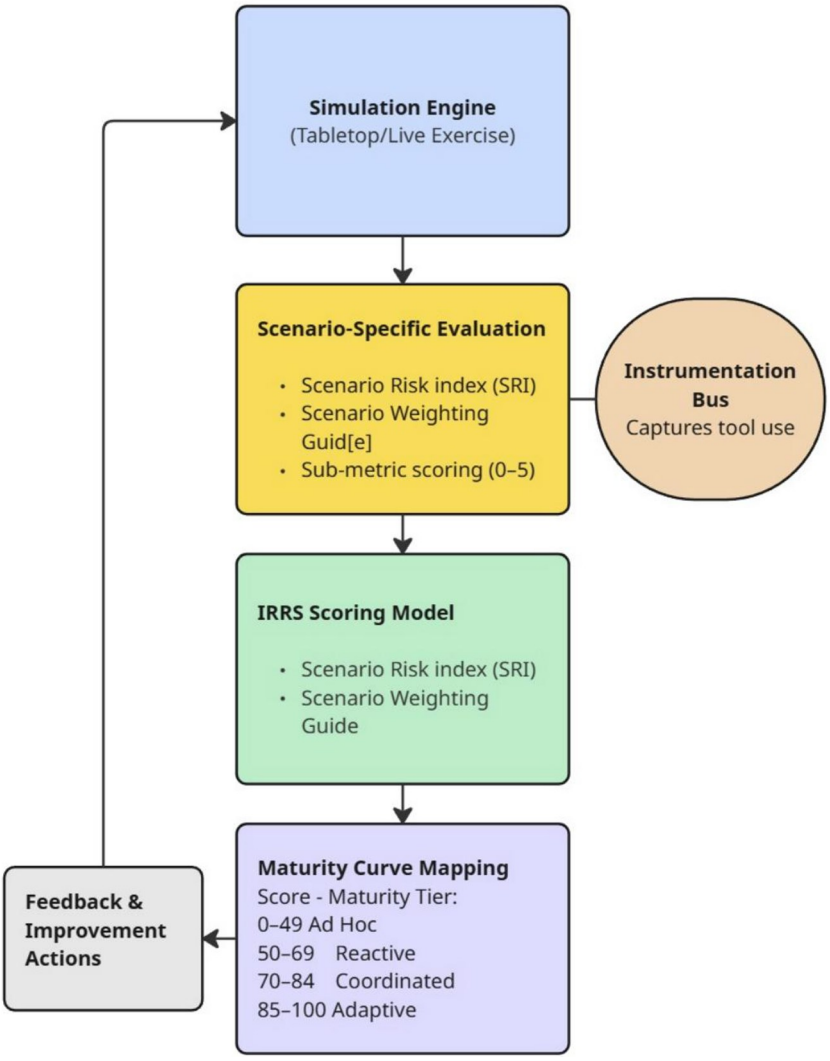

Figure 1. The IRRS framework architecture and flow
Source: Authors' own work

capture throughout, supporting evidence-based assessment. A feedback loop ensures simulation outcomes directly inform iterative improvement efforts. This end-to-end flow reinforces the IRRS model's emphasis on contextual realism, structured evaluation and continuous organisational uplift.

4.1 Simulation engine

The simulation engine generates structured, threat-informed scenarios that replicate the pressure, ambiguity and coordination demands of real-world incidents, enabling empirical observation of response behaviours. To accommodate varying maturity levels, the engine offers two modes, as summarised in Table 2. Tabletop simulations assess procedural alignment and decision-making in a risk-free setting, making them suitable for early-stage organisations. Conversely, live emulations provide high-fidelity testing through dynamic telemetry and real-time containment tasks, validating technical controls in more mature environments. Delivered through a unified simulation framework, this dual-mode approach ensures scalability and contextual relevance across diverse organisational profiles.

The simulation engine supports multiple scenario types, each designed to emulate threat conditions prevalent within SMEs. These scenarios enable organisations to assess targeted aspects of their IR capability under varying levels of risk and complexity. Selection is informed by organisational threat models and tailored to align with sectoral priorities and exposure levels.

4.2 Scenario-specific scoring and the Incident Response Readiness Score model

Following simulation, IRRS applies scenario-specific scoring to quantify performance relative to risk profile and organisational context. The framework uses diverse scenario archetypes selected for their prevalence and impact in SME environments to test distinct response capabilities. Currently, four scenarios are supported: phishing-led ransomware, insider data exfiltration, public-cloud misconfiguration and credential leakage (Table 3).

The SRI is a calibrated ordinal scale used within the IRRS framework to weight simulation outcomes according to the operational and strategic significance of the scenario under evaluation. It ensures that readiness assessments reflect not only how well a task was performed, but also how critical that task was within a particular threat context. This approach aligns with risk-adjusted evaluation principles widely adopted in vulnerability scoring systems (e.g. CVSS) and enterprise risk management practices. Each scenario is assigned an SRI value ranging from 1 (very low risk) to 5 (very high risk), determined through structured pre-simulation risk modelling.

Table 2. Simulation modes and scenarios

Mode	Function
Tabletop simulation	Evaluates procedural alignment, strategic decision-making and team coordination in a risk-free environment. Key artefacts include scenario playbooks, timed inject schedules, facilitator scripts and observation checklists
Live emulation	Tests technical detection, containment and eradication capabilities within isolated virtualised or cloud environments. Artefacts include virtual machines, Infrastructure-as-Code templates, seeded log data and adversary emulation scripts aligned with MITRE ATT&CK

Source(s): Authors' own work

Table 3. Scenario types

Scenario type	Details
Phishing-led ransomware propagation	Simulates a phishing compromise followed by lateral movement and ransomware deployment, resulting in data encryption and extortion. Primary risk vectors include social engineering and endpoint compromise
Insider data exfiltration	Models an insider using removable media or personal cloud storage to exfiltrate sensitive data. Risk vectors include insider misuse and unauthorised data access
Public-cloud misconfiguration	Emulates exposure of personally identifiable information (PII) because of misconfigured access controls, such as open AWS S3 buckets. Risk vectors include misconfiguration and poor cloud governance
Credential leakage	Replicates unauthorised access via credentials exposed in public code repositories or compromised CI/CD pipelines. Risk vectors include credential theft and supply chain exposure

Source(s): Authors' own work

This five-level scale, as presented in Table 4, mirrors severity models commonly adopted in operational risk, threat modelling frameworks such as STRIDE and incident management maturity models. It is deliberately designed as an ordinal scale rather than a ratio-based one for several reasons. First, it emphasises the relative business impact of an incident rather than requiring precise quantification of absolute loss, which is often impractical for SMEs because of limited data and resources. Second, it reflects the triage-based categorisation practices prevalent in many SME environments, where structured classification is more feasible than exhaustive impact measurement. Third, it supports interoperability with detection pipelines, structured playbooks and prioritisation models used in IR planning and SOC workflows.

Table 4. IRRS SRI levels

SRI level	Risk description and representative scenario
SRI 1 – very low risk	Operational impact is negligible. Effects are isolated, non-sensitive and quickly reversible, with no strategic consequences. Example: benign adware detected; no sensitive data accessed
SRI 2 – low risk	Localised impact with limited propagation. May reveal hygiene issues but poses no critical threat. Example: misconfigured antivirus suppresses alerts, and minor malware is quarantined without incident
SRI 3 – moderate risk	A risk of escalation or lateral movement exists. Impact is significant if mishandled, though not immediately urgent. Example: a phishing link captures privileged credentials, but the account remains unused
SRI 4 – high risk	Causes business disruption or data exposure. Requires a timely and coordinated response to mitigate legal or reputational consequences. Example: insider exports sensitive customer data via USB
SRI 5 – very high risk	Catastrophic scenario involving widespread compromise, system unavailability or organisational viability risk. Example: ransomware spreads across core servers with exfiltration and extortion

Source(s): Authors' own work

Within the IRRS framework, each SRI value functions as a weight multiplier, amplifying or moderating the influence of simulation performance depending on scenario severity. Consequently, failures in high-risk simulations (SRI 4–5) are scored more stringently than those in lower risk contexts (SRI 1–2). This proportional scoring approach not only ensures fairness but also incentivises improved readiness where the potential for harm is greatest. Furthermore, all IRRS simulations are evaluated against 15 immutable sub-metrics grouped into five readiness domains (Table 5). Fixing the metric catalogue ensures longitudinal comparability across time, teams and organisations. These sub-metrics form the backbone of IRRS evaluation, ensuring consistency and comparability across scenarios.

During a simulation, each of the 15 sub-metrics shown in Table 5 is assessed by a designated evaluator based on observed team behaviour. These scores, ranging from 0 (not met) to 5 (fully met), are then combined with scenario-specific weights defined by the SRI to produce a normalised readiness score. This process enables consistent, risk-adjusted evaluation across diverse incident types.

4.3 Incident Response Readiness Score scoring formula

The IRRS scoring process consists of three sequential stages that transform observed behaviours during the simulation into a scenario-weighted, normalised readiness score. The formula is structured to explicitly reflect the integration of evaluator assessments and scenario-derived risk weighting, with the formula symbols and interpretive notes defined in Table 6:

$$\mathrm{IRRS} = 100 \times \frac{\sum_{i=1}^{n} (s_i \times \omega_i)}{5 \times \sum_{i=1}^{n} \omega_i}$$

Computation proceeds in three steps:

(1) Each sub-metric score s_i is multiplied by its scenario w_i.

(2) The weighted scores are summed to yield a raw total.

Table 5. IRRS sub-metrics

Readiness domain	Sub-metrics
Procedural alignment	• Escalation path followed • IRP referenced during incident • Deviations from IRP formally justified
Operational execution	• Containment-action timing • Task coverage (breadth of technical actions) • Execution accuracy
Infrastructure integration	• Tool-usage effectiveness • Tool alignment to IRP • Inter-tool visibility
Coordination and communication	• Role clarity • Decision flow • Communication logging
Post-incident follow-through	• Root-cause analysis • Lessons learned • IRP updated post-simulation

Source(s): Authors' own work

Table 6. IRRS formula analysis

Symbol	Definition	Notes
n	Total number of universal sub-metrics (fixed at 15)	Ensures every scenario is scored on the same metric set
s_i	Evaluator score for sub-metric i	0 = not met, 5 = fully met (ordinal scale)
w_i	Scenario-specific weight for sub-metric i	1–5, set by the Scenario Risk Index (SRI) table
$5\sum w_i$	Maximum raw score attainable in this scenario	Multiplies the perfect evaluator mark (5) by the cumulative risk weights; dynamically adjusts to any scenario

Source(s): Authors' own work

(3) The raw total is normalised by the maximum attainable score $5\sum w_i$ and scaled to a [0–100] range.

This formulation guarantees cross-scenario comparability. Heavily weighted, high-risk exercises enlarge the denominator in proportion to their severity, so performance is always reported as a true percentage of "risk-adjusted perfection".

4.4 Scenario weighting guide and evaluator scoring rubric

To enhance reproducibility and scoring consistency across facilitators, we have developed two supplementary resources available via GitHub (Abid, 2025):

Table 7 provides a sample of recommended weight values (1–5) for selected IRRS sub-metrics across common scenarios (e.g. ransomware and insider threats). Each weight is justified based on its criticality and role within the incident type.

Table 8 offers structured criteria to guide evaluators when scoring organisational behaviour during simulations. Each sub-metric includes behavioural anchors for scoring from 0 to 5.

To further strengthen reproducibility and evaluation consistency, future iterations of the IRRS framework should include controlled benchmarking exercises involving multiple independent evaluators scoring the same simulation. This would enable measurement of inter-rater reliability and help calibrate scoring expectations across varied facilitator backgrounds. While structured scoring rubrics and scenario weighting guides have been developed to minimise ambiguity, such validation studies would empirically test the framework's ability to support consistent scoring in realistic, high-pressure settings. Early pilot designs could include cross-sectional scoring of recorded simulations or "double-blind" facilitator exercises within the same SME environment.

Table 7. Sample excerpt from scenario weighting guide

Scenario type	Sub-metric	Weight	Justification
Ransomware	Operational execution – task coverage	5	Broad containment across multiple systems is critical to halting lateral spread
Insider threat	Operational execution – task coverage	4	Focus may shift to identity-specific controls and data access channels

Source(s): Authors' own work

Table 8. Sample excerpt from evaluator scoring rubric

Scenario type	Score = 0	Score = 3	Score = 5
Task coverage (ransomware)	Only one host isolated	Most compromised hosts were isolated; some lateral spread was contained	Full containment of all compromised assets
Root-cause analysis (credential leak)	No RCA performed	Immediate cause identified, but systemic control weaknesses only partially examined	Multi-layered RCA covering process and tooling

Source(s): Authors' own work

4.5 Incident Response Readiness Score worked examples

To illustrate how the risk-adjusted formula behaves under different threat contexts, two complete calculations are presented in Tables 9 and 10: insider-threat and ransomware propagation. The weights come from the framework's scenario matrices; the evaluator scores (s_i) are the scores given by the evaluator during the exercise on the 0–5 scale. The $s_i w_i$ is the multiplication of the weight by the evaluator score:

$$\text{IRRS} = 100 \times \frac{183}{5 \times 60} = 61.0$$

$$\text{IRRS} = 100 \times \frac{198}{5 \times 57} \approx 69.5$$

The results illustrate how risk normalisation within the IRRS formula functions effectively. Although the ransomware scenario carries a slightly lower total weight ($\Sigma\ w_i = 57$) compared

Table 9. Insider threat scenario

Domain	Sub-metric	s_i	w_i	$s_i w_i$
Procedural alignment	Escalation path followed	4	5	20
	IRP referenced during incident	3	5	15
	Deviations justified	4	4	16
Operational execution	Containment-action timing	3	4	12
	Task coverage	2	4	8
	Execution accuracy	4	5	20
Infrastructure integration	Tool-usage effectiveness	3	4	12
	Tool alignment to IRP	2	3	6
	Inter-tool visibility	3	3	9
Coordination and comms	Role clarity	4	5	20
	Decision flow	3	5	15
	Communication logging	3	4	12
Post-incident follow-through	Root-cause analysis	3	3	9
	Lessons learned	2	3	6
	IRP updated post-simulation	1	3	3
Totals			$\Sigma w_i = 60$	$\Sigma\ s_i w_i = 183$

Source(s): Authors' own work

Table 10. Ransomware propagation scenario

Domain	Sub-metric	s_i	w_i	$s_i w_i$
Procedural alignment	Escalation path followed	4	4	16
	IRP referenced during incident	4	4	16
	Deviations justified	3	3	9
Operational execution	Containment-action timing	4	5	20
	Task coverage	4	5	20
	Execution accuracy	4	5	20
Infrastructure integration	Tool-usage effectiveness	3	4	12
	Tool alignment to IRP	3	3	9
	Inter-tool visibility	3	3	9
Coordination and comms	Role clarity	4	4	16
	Decision flow	4	5	20
	Communication logging	3	4	12
Post-incident follow-through	Root-cause analysis	3	3	9
	Lessons learned	2	3	6
	IRP updated post-simulation	2	2	4
Totals			$\Sigma\, w_i = 57$	$\Sigma\, s_i w_i = 198$

Source(s): Authors' own work

to the insider-threat scenario ($\Sigma w_i = 60$), the denominator in the IRRS equation scales proportionally, ensuring both outcomes are directly comparable on the same 0–100 scale. The difference in performance between the two simulations is also informative. The team achieved a readiness score of 61% in the insider-threat exercise and approximately 69.5% in the ransomware simulation. This suggests a comparatively stronger capability in responding to high-intensity, time-critical incidents, particularly where rapid containment and technical execution are essential. Furthermore, the insider-threat score was notably affected by underperformance in task coverage and post-simulation IRP updates, two sub-metrics with weights of 4 and 3, respectively. In spite of their moderate weighting, deficiencies in these areas had a measurable impact on the overall readiness score. This underscores that improving lower scoring, moderately weighted behaviours may yield more significant readiness gains than focusing on marginal improvements in already well-performing sub-metrics.

Together, these comparative calculations validate the IRRS scoring model's fairness and diagnostic utility. It accommodates threat-specific weighting while still producing a unified, interpretable metric that supports benchmarking across scenarios and informs targeted organisational improvement.

5. Incident Response Readiness Score maturity curve classification

The maturity curve serves as the framework's interpretive layer, translating normalised scores into tiered qualitative assessments and supporting executive decision-making. It fulfils three functions: enabling standardised benchmarking, abstracting complex metrics for non-technical stakeholders and defining a clear developmental trajectory. The tiering logic is grounded in empirically observed behavioural patterns from simulations, establishing ordinal stages that reflect progressive improvements in procedural adherence, coordination, tooling integration and adaptive capacity.

The score ranges were selected based on threshold effects observed in simulation scoring patterns. Organisations scoring below 50 typically demonstrate inconsistent role adherence,

informal escalation paths and unstructured decision-making, indicating that IR efforts are largely improvised. In the 50–69 range, procedural elements may be present but are applied inconsistently, often depending on individual effort rather than coordinated team execution. Scores between 70 and 84 reflect maturing capabilities, with most response actions mapped to defined processes and moderate tooling use, though continuous improvement and integration remain limited. Scores of 85 and above reflect highly structured, proactive and context-driven response capabilities, supported by regular post-incident learning and real-time decision-making.

This tiering structure ensures that readiness classifications are behaviourally meaningful and proportionally aligned to simulation severity and sub-metric weightings. To assign a maturity tier, the IRRS score is first calculated and normalised to a [0–100] scale. The result is mapped to a tier using the fixed bands shown in Table 11. For scores within ±2 points of a boundary (e.g. 68–72), evaluators may apply professional discretion based on observed behaviours and qualitative performance notes. Final classifications are recorded alongside domain-specific feedback and recommended uplift actions.

6. Incident Response Readiness Score instrumentation bus

A key design challenge in scenario-based readiness assessment is ensuring that evaluator scoring reflects observable reality rather than post hoc rationalisation or facilitator bias. The IRRS framework addresses this challenge through the integration of an instrumentation bus, a structured, passive observation mechanism required to be built in or factored into the simulation environment to capture relevant human and system interactions without influencing team behaviour. The objective of this component is to ensure that each sub-metric score is grounded in verifiable evidence relating to procedural alignment, tool invocation, escalation timing and communication hand-offs. The instrumentation bus is not a rigid component but a flexible design layer that must be planned in accordance with simulation scope, available infrastructure and organisational maturity. In low-technology environments, this may be implemented entirely through facilitator forms and structured observation. In more advanced live emulation contexts, it may include telemetry capture from virtual machines, adversary emulation platforms or SIEM pipelines. What is essential is not the specific instrumentation mechanism but the deliberate incorporation of observational fidelity into simulation design. Without such planning, IRRS scores risk becoming overly dependent on facilitator inference or participant interpretation. The bus ensures that metrics related to timing, coordination and procedural adherence are based on verifiable events, thereby strengthening the objectivity, repeatability and defensibility of readiness assessments. In summary, the instrumentation bus is critical for aligning simulation-based assessment with IRRS objectives. It enables scoring to reflect not just whether key actions were taken but how, when and by whom? These dimensions are fundamental to understanding the true operational readiness of an organisation under incident pressure.

Table 11. IRRS maturity tier mapping

IRRS score range	Maturity tier
0–49	Ad hoc
50–69	Reactive
70–84	Coordinated
85–100	Adaptive

Source(s): Authors' own work

7. Human-centric telemetry and cognitive process metrics

The instrumentation bus described in the preceding section establishes the evidentiary backbone of the IRRS framework by ensuring that evaluator judgements are grounded in observable events rather than retrospective interpretation. While this mechanism is sufficient for validating what actions were taken, it does not, on its own, explain why those actions occurred when they did, nor why specific response paths emerged under pressure. This distinction is particularly important in SME environments, where IR outcomes are often shaped as much by human decision-making constraints as by technical capability.

Existing IRRS scoring domains intentionally focus on outcomes and execution fidelity, for example, whether containment was achieved, whether escalation followed defined procedures and whether post-incident learning occurred. However, feedback from both practitioners and reviewers of the conference version of this work highlighted a recurring limitation: identical IRRS scores may arise from fundamentally different underlying organisational dynamics. A delayed containment action may reflect inadequate tooling in one organisation but cognitive overload, unclear authority boundaries or coordination friction in another. To address this explanatory gap, this section extends the instrumentation bus with a set of human-centric telemetry metrics designed to capture process-level friction during IR simulations. These metrics do not attempt to assess individual competence, intent or psychological traits. Instead, they quantify systemic interactions between human operators, organisational structure and technical interfaces, as manifested through timing, communication patterns and role adherence. The objective is not attribution of blame, but diagnostic clarity.

More importantly, these metrics are designed as diagnostic complements to IRRS rather than as independent maturity indicators. They do not alter the IRRS score itself; instead, they provide structured explanations for why specific scores emerged, thereby strengthening the framework's analytical depth and practical utility.

7.1 Design principles

The human-centric metrics introduced in this section were designed to provide explanatory insight into IR performance while remaining practical, reproducible and ethically neutral. Each metric is derived exclusively from artefacts already captured by the IRRS instrumentation bus, such as timestamps, communication records and structured evaluator observations, ensuring that measurement does not depend on intrusive monitoring or post hoc participant self-reporting. The metrics are intentionally framed at the system level, capturing organisational and procedural dynamics rather than individual cognitive performance. Observed delays, coordination overhead or deviations from expected behaviour are interpreted as properties of the response system itself, including governance structures, decision pathways and tool integration, rather than as individual failings. To support consistent application, the metrics are designed to be scenario-agnostic and independent of organisational scale, allowing them to be applied across different incident types and simulation modes without redefinition. Finally, the formulations prioritise analytical parsimony, relying on simple interval- and ratio-based measures that can be captured reliably by evaluators under time pressure, thereby reducing ambiguity and scoring variance.

The selection of decision latency, communication entropy and authority drift as the three human-centric telemetry metrics was directly motivated by the event types and artefacts captured by the IRRS instrumentation bus. Decision latency is derived from inject timestamps and action timestamps, artefacts that the instrumentation bus records by design to support submetric scoring of containment timing and escalation adherence. Communication entropy is computed from communication transcripts and valid action counts, both of which are standard instrumentation bus outputs during high-stress phases such as triage and

containment. Authority drift is calculated from role scope declarations and action logs, artefacts that the bus collects to validate role clarity and decision flow submetrics. In other words, each metric was defined because the instrumentation bus already produces the raw evidence required to compute it without any additional instrumentation overhead. Other candidate metrics, such as absolute alert volume or individual task completion rate, were considered but excluded because they either required intrusive monitoring beyond the passive design of the bus or conflated technical detection performance (already captured in IRRS scoring) with human system friction (the intended focus of the telemetry layer). Collectively, these three metrics were selected as a minimal, non-overlapping set that captures temporal delay (decision latency), coordination inefficiency (communication entropy) and structural breakdown (authority drift), thereby providing orthogonal explanatory coverage of human system interaction during IR.

Guided by these considerations, three complementary metrics were defined: decision latency, communication entropy and authority drift. Together, they capture distinct but interrelated dimensions of human–system interaction during IR, providing diagnostic context for IRRS outcomes without altering the core scoring model.

7.2 Measurement scope and instrumentation bias

To ensure methodological clarity and reproducibility, each human-centric telemetry metric defined in this section is anchored to a specific unit of observation, applied at a defined level of organisational abstraction and derived exclusively from artefacts available to the instrumentation bus. These metrics are event-based rather than continuous. They do not attempt to model general cognitive performance across the entire simulation; instead, they quantify organisational behaviour at decision-critical moments that materially influence incident outcomes. This design choice reflects both practical evaluator constraints and established IR practice, where only a subset of alerts or actions meaningfully shape response trajectory.

Specifically, the scope of each metric is defined as follows:

Decision latency is measured per decision-critical inject – that is, injects that present actionable intelligence requiring a non-trivial response decision such as containment, escalation or credential revocation. Not all alerts are included; only those that could reasonably alter the course of the incident are instrumented.

Communication entropy is measured per high-stress response phase, such as triage or containment, during which coordination demands and time pressure are highest. Routine briefings and post-incident discussions are excluded to avoid diluting phase-specific signals.

Authority drift is measured per critical action, defined as actions that materially affect incident trajectory, governance posture, or organisational risk exposure. Minor or purely mechanical actions are excluded.

All measurements are conducted at the system level, not the individual level. Timing, communication volume and role adherence are evaluated as properties of organisational processes rather than personal performance. This boundary is intentional and ensures that the metrics diagnose structural and procedural friction without attributing fault to individual actors. Authoritative data sources for all three metrics are restricted to verifiable artefacts already captured by the instrumentation bus, including alert timestamps, tool invocation logs, facilitator inject records, communication transcripts and pre-declared role definitions. No post hoc participant self-reporting or inference of intent is used in metric calculation. By explicitly defining measurement scope, granularity and evidentiary sources, this framework ensures that human-centric telemetry remains reproducible, non-intrusive and analytically defensible across diverse contexts.

7.3 Decision latency

Decision Latency captures the temporal separation between information availability and decisive action, isolating human cognitive processing time from technical detection latency. In SME environments, delays are frequently attributed to inadequate tooling or alerting; however, empirical observation during simulations suggests that delays often persist even when actionable intelligence is clearly visible and technically actionable. Decision latency is defined as the elapsed time between the moment actionable incident information becomes visible to a responsible operator and the moment a valid containment or response command is executed:

$$D_L = T_{action} - T_{visible}$$

where:

$T_{visible}$ = timestamp at which actionable intelligence is presented to the operator; and

T_{action} = timestamp of the first valid technical action issued in response.

A high D_L value relative to scenario complexity indicates elevated cognitive load, uncertainty or organisational hesitation. Crucially, this interpretation holds regardless of whether the eventual action is correct. In several observed cases, extended latency was associated with behaviours such as excessive verification, fear of business disruption from false positives or uncertainty regarding escalation authority. As such, decision latency functions as a proxy indicator for decision confidence and psychological safety within the response system.

Within the instrumentation bus, decision latency is captured via the inject timestamp recorded by the facilitator at the moment actionable intelligence is delivered and the corresponding action timestamp logged when the first valid technical command is issued. Full recording guidance, including valid action rules, threshold calibration tables and a worked example, is provided in the Decision Latency Capture Sheet (Checklist A) available in the public repository (Abid, 2025).

7.4 Communication entropy

Effective IR requires communication, but excessive or unstructured communication can become a source of friction rather than coordination. Communication entropy operationalises this phenomenon by measuring the signal-to-action ratio within response communications. Communication entropy is defined as the ratio of discrete communication artefacts (e.g. chat messages, verbal exchanges) to successful technical actions executed during a defined response phase:

$$C_E = \frac{\sum M_{total}}{A_{valid}}$$

where:

M_{total} = total count of discrete messages or utterances; and

A_{valid} = number of successful technical response actions.

Higher ratios indicate increasing coordination overhead. Empirical thresholds observed during simulations suggest that ratios exceeding approximately 20:1 are characteristic of "storming" dynamics, where teams seek consensus or permission rather than executing predefined protocols. Importantly, this metric does not penalise collaboration; instead, it

highlights the absence of pre-authorised decision triggers and role clarity, which force teams to negotiate each step in real time.

Communication entropy is computed from communication transcripts and valid action logs produced by the instrumentation bus during designated response phases. Full recording guidance, including counting rules, communication pattern classification, threshold calibration tables and a worked example, is provided in the Communication Entropy Tally Sheet (Checklist B) available in the public repository (Abid, 2025).

7.5 Authority drift

Authority drift captures the extent to which organisational role boundaries erode under incident pressure. In SMEs, it is common for senior leaders to intervene in technical tasks ("downwards drift") or for technical staff to assume strategic decision-making responsibilities ("upwards drift"), often with unintended consequences. Authority drift is defined as the proportion of critical actions taken by actors outside their predefined role scope:

$$A_D = \left(\frac{A_{\text{out of scope}}}{A_{\text{total}}} \right) \times 100$$

where:

$A_{\text{out of scope}}$ = represents actions taken outside designated role boundaries; and
A_{total} = total number of critical actions observed.

High authority drift indicates brittle governance structures in which formal response plans are abandoned in favour of individual improvisation. While such behaviour may occasionally accelerate action, it frequently undermines accountability, disrupts coordination and leaves strategic risk unmanaged. Within the IRRS context, authority drift provides a structural explanation for low scores in role clarity and decision flow sub-metrics.

Authority drift is derived from pre-declared role definitions and action logs captured by the instrumentation bus. Full recording guidance, including the mandatory pre-simulation role registry, drift type classification, threshold calibration tables and a worked example, is provided in the Authority Drift Tracker (Checklist C) available in the public repository (Abid, 2025).

To illustrate what constitutes an in-scope versus out-of-scope action during the Insider Threat scenario, consider the following examples drawn from the worked example in the Authority Drift Tracker instrument (Checklist C) available in the public repository (Abid, 2025). In-scope actions for the IT Security Analyst role include disabling a user session upon detection of anomalous file access, applying a USB restriction following confirmation of removable media usage, revoking a cloud access token upon evidence of exfiltration and raising a formal escalation to the incident commander per the IRP, each falling within pre-authorised response boundaries. By contrast, two out-of-scope actions were recorded during the worked example. The first was an instance of downwards drift, in which a senior manager directly accessed and modified DLP alert configuration settings, a technical action reserved for the analyst role, causing a configuration inconsistency and temporary loss of alert visibility. The second was an instance of upwards drift, in which a technical staff member independently escalated the incident to external legal counsel without managerial authorisation, a governance action outside the technical role boundary that created external disclosure risk without executive sign-off. These two drift types, downwards and upwards,

have distinct governance remediation implications and are reported separately in the Authority Drift Tracker.

7.6 Relationship to Incident Response Readiness Score and evaluator instrumentation

Human-centric telemetry metrics serve as a diagnostic complement to the IRRS, contextualising readiness outcomes without altering domain scores or maturity tiers. By surfacing latent human–system dynamics – such as cognitive overload, coordination friction or ambiguous authority – these metrics explain why identical readiness scores may arise from distinct organisational conditions, shifting evaluation from performance measurement towards diagnostic attribution.

Full instrumentation resources, including the Decision Latency Capture Sheet, Communication Entropy Tally Sheet and Authority Drift Tracker, are available via the IRRS public repository (Abid, 2025), which has been updated to version 2.0 to include these human-centric telemetry instruments alongside the existing scenario weighting guides and evaluator rubrics. Each instrument includes a pre-simulation setup section, structured recording tables, a communication pattern or drift type classification guide and a fully worked example aligned with the insider threat scenario used throughout this paper. Critically, each instrument also includes scenario-contextualised threshold guidance specifying when each metric should be considered acceptable, elevated or critical relative to the scenario class and SRI level being evaluated. For example, a decision latency exceeding 10 min is classified as critical in SRI 5 scenarios such as active ransomware propagation, where rapid containment is paramount, while the same duration falls only within the elevated band during an insider threat investigation where deliberate verification before action is operationally appropriate. Acceptable ranges vary by scenario class across all three instruments, and evaluators are directed to the repository for the full calibration tables associated with each.

7.7 Human-centric telemetry worked examples and interpretation bands

To illustrate how the human-centric telemetry metrics defined earlier are applied in practice, this section presents a single, consolidated theoretical worked example based on an insider threat scenario aligned with the corresponding IRRS worked example. The purpose of this worked example is to demonstrate the calculation mechanics, presentation format and interpretive value of decision latency, communication entropy and authority drift when used as diagnostic complements to IRRS. The telemetry values presented in this section are illustrative rather than empirical, with the decision latency values presented in Table 12 constructed to reflect the types of artefacts the IRRS instrumentation bus is designed to capture (e.g. inject timestamps, action timestamps, communication logs and role-scope declarations). This approach is adopted to transparently demonstrate the method without over-claiming empirical validation; further empirical testing, repeated simulation runs and evaluator cross-scoring are required to formally validate expected telemetry ranges and inter-rater reliability.

- Decision latency:

$$\bar{D}_L = \frac{7 + 12 + 14}{3} = 11.0\,\text{min},$$

with a maximum observed latency of 14 min.

Table 12. Insider threat decision latency data

Inject	Decision-critical event	Decision latency (Min)	First valid response
1	Unusual file access spike detected	7	Disable user session
2	Evidence of removable media usage	12	Apply temporary USB restrictions
3	Confirmed data copied to personal cloud	14	Revoke token and escalate

Source(s): Authors' own work

- Communication entropy

The communication entropy values used for each phase are presented in Table 13 and calculated as:

$$C_E^{triage} = \frac{46}{2} = 23.0$$

$$C_E^{containment} = \frac{78}{3} = 26.0$$

Resulting in a communication entropy range of 23.0–26.0, indicating elevated coordination overhead during early response phases.

- Authority drift:

The authority drift values are presented in Table 14 and calculated as:

$$A_D = \left(\frac{3}{10}\right) \times 100 = 30\%$$

This indicates that nearly one-third of critical actions were taken outside predefined role boundaries during the scenario.

(2) Human-centric summary and interpretation

Table 13. Insider threat communication entropy data

Phase	Messages	Valid actions
Triage	46	2
Containment	78	3

Source(s): Authors' own work

Table 14. Insider threat authority drift data

Quantity	Count
Total critical actions	10
Out-of-scope actions	3

Source(s): Authors' own work

Human-centric telemetry serves as an interpretive layer, explaining why readiness outcomes emerged under operational pressure. Analysis indicates that response effectiveness was constrained less by technical capability than by decision confidence and coordination efficiency. Elevated decision latency signals cognitive load and risk aversion, often stemming from unclear authority triggers or verification loops. High communication entropy reflects inefficient consensus-seeking, where increased communication volume fails to translate into decisive technical action. Finally, authority drift exposes governance fragility, as out-of-scope interventions dilute accountability. Collectively, these metrics distinguish technical limitations from human–system friction, enabling diagnostic attribution and targeted remediation without altering core IRRS classification.

Table 15 summarises the three telemetry dimensions across both scenarios, highlighting consistent patterns whereby elevated latency and authority drift are associated with lower IRRS classifications, while higher scoring scenarios exhibit more decisive, role-adherent response behaviour. To connect these results directly to the IRRS worked example in Section 5, the insider threat scenario's IRRS of 61 (reactive tier) is explained by its telemetry profile. The mean decision latency of 11 min, peaking at 14 min on the confirmed exfiltration inject, corresponds to the low score in containment-action timing ($si = 3$, $wi = 4$), reflecting hesitation before the token revocation decision. The communication entropy range of 23–26 explains moderate scores in communication logging and decision flow, as high message volumes relative to valid actions reflect a team seeking consensus rather than executing pre-authorised protocols. The 30% authority drift is the primary diagnostic explanation for low performance in role clarity and decision flow, as out-of-scope interventions disrupted the established response chain. By contrast, the ransomware scenario's IRRS of 69.5 is consistent with substantially lower telemetry values, specifically a mean decision latency of 5.5 min, communication entropy of 9–11 and authority drift of 7.1%, reflecting a more decisive, role-adherent response under high-intensity conditions where containment authority was clearer and pre-authorised actions were more readily invoked.

8. Feedback and improvement actions

The IRRS framework incorporates a structured feedback mechanism to ensure that insights derived from simulations are translated into actionable improvements. Post-simulation reviews synthesise sub-metric outcomes, behavioural observations and instrumentation data to identify specific areas of strength and weakness. These findings inform targeted remediation – such as refining escalation paths, enhancing tool integration, or clarifying response roles. Iterative testing using updated scenarios allows organisations to measure progress over time and validate the impact of interventions. This feedback loop reinforces IRRS's core objective: enabling continuous uplift of IR capability through evidence-based learning. Over successive simulation cycles, tracked improvements can elevate an organisation's IRRS maturity tier, enabling strategic alignment between operational uplift and measurable readiness progression.

Table 15. Human-centric telemetry summary

Scenario	IRRS	$\bar{D}_L$	D_L^{max}	C_E	A_D (%)
Insider threat	61	11	14	23–26	30
Ransomware	69.5	5.5	8	9–11	7.1

Source(s): Authors' own work

9. Limitations and bias mitigation

While the IRRS framework offers a structured and practical method for evaluating SME IR readiness, certain limitations and potential sources of bias should be acknowledged. One notable source of potential bias lies in evaluator subjectivity. Although structured rubrics and scenario weighting guides are provided, scoring outcomes may still be influenced by the evaluator's interpretation, experience or professional judgement. Differences in how sub-metrics or scenario-critical actions are perceived can result in inconsistent application of scores across evaluators. This highlights the need for future benchmarking studies involving multiple evaluators to assess inter-rater reliability and improve scoring consistency.

In addition, the human-centric telemetry metrics introduced in this study introduce their own measurement considerations. While these metrics are derived from artefacts already available to the IRRS instrumentation bus, their interpretation relies on structured observation of decision timing, communication patterns and role adherence during simulations. As such, the telemetry values presented in this work are illustrative and theoretical, intended to demonstrate diagnostic feasibility rather than establish empirically validated thresholds. Future work will, therefore, focus on assessing inter-evaluator consistency, validating expected telemetry ranges across repeated simulations and examining the stability of these metrics across organisational contexts.

It should further be noted that the instrumentation documents available in the public repository (Abid, 2025) have been substantially revised in version 2.0 to address the inherently context-dependent nature of metric criticality. Each instrument now includes scenario contextualised threshold guidance specifying acceptable, elevated and critical bands by scenario class and SRI level, as well as fully worked examples demonstrating how recorded values are interpreted and classified. Of particular note, the Decision Latency Capture Sheet clarifies that a 10-min latency carries materially different implications in a ransomware scenario requiring immediate containment than in a phishing-driven insider threat scenario where premature action may compromise the investigation. Similarly, the Authority Drift Tracker now includes a mandatory pre-simulation Role Registry, without which drift cannot be reliably assessed and distinguishes between downwards drift (strategic roles performing technical actions) and upwards drift (technical roles assuming executive decisions), as each pattern has distinct governance remediation implications. These thresholds and classifications are presented as practitioner-informed reference points to be refined and empirically validated through future simulation cycles rather than fixed normative standards.

The framework's scope is also limited by scenario coverage. IRRS currently focuses on four representative incident types – ransomware propagation, insider data exfiltration, public-cloud misconfiguration and credential leakage. While these scenarios reflect prevalent SME threats, they do not cover the full breadth of emerging attack vectors such as supply chain compromise, advanced persistent threats or zero-day vulnerabilities. As such, findings may not fully generalise to all organisational contexts.

Another risk involves the potential for false confidence. Because the IRRS produces a quantified score and mapped maturity tier, organisations may misinterpret strong performance in one simulation as a proxy for comprehensive readiness. This may result in underappreciation of vulnerabilities not captured in the scenario design. It is, therefore, recommended that organisations conduct regular, varied simulation cycles and periodically update scenarios to maintain realism and mitigate overconfidence.

Finally, the method may under-represent latent strengths or weaknesses that do not manifest during a given simulation. Simulation outcomes reflect observable behaviours under controlled conditions, which may not reveal all relevant organisational dynamics.

Important capabilities – such as leadership initiative under pressure, undocumented escalation paths or communication breakdowns – could go unnoticed without broader qualitative inquiry. Repeated simulations, triangulated with post-exercise debriefs and observational data, may help surface these less visible factors. These limitations do not undermine the value of the IRRS framework but rather define a roadmap for future research. Specifically, multi-evaluator benchmarking studies, expansion of scenario libraries and longitudinal tracking of simulation outcomes across diverse SME environments represent logical next steps in advancing IRRS maturity and reproducibility. The GitHub material (Abid, 2025) provides a foundation for future enhancement and validation of the framework, supporting both customisation and replication to mitigate the limitations identified.

10. Discussion

This study has introduced and evaluated the IRRS, a scenario-driven framework designed to empirically quantify the operational readiness of SMEs. By integrating structured simulations with a risk-weighted scoring model, IRRS moves beyond traditional compliance-based assessment to capture the behavioural, procedural and technical dimensions of IR capability.

The baseline scenario evaluations demonstrate that critical gaps persist in SMEs' ability to operationalise their IRPs. In spite of the presence of documented procedures, evaluations using this baseline revealed patterns of *ad hoc* decision-making, delayed containment and informal escalation. These findings align with broader industry observations regarding the static and performative nature of many SME security policies. In addition, underutilisation of existing tools and the absence of structured post-incident documentation limited the organisation's maturity classification to the reactive tier. The inclusion of human-centric telemetry provided explanatory depth by revealing how cognitive and organisational dynamics shaped response effectiveness. Elevated decision latency following the availability of actionable intelligence, together with authority drift during containment, explained why documented procedures failed to translate into timely execution. These results show that IRRS scores capture the level of readiness achieved, while human-centric telemetry clarifies the underlying conditions that produced that outcome under operational pressure. Application of the framework further confirmed that these shortcomings are not anecdotal but can be systematically observed and scored. Passive observation through the instrumentation bus enabled real-time assessment of containment timeliness and role clarity without disrupting the simulation. This ensured that evaluation remained grounded in observable evidence rather than subjective interpretation, establishing the instrumentation bus as a critical design element for future IRRS-aligned simulations.

From a practical perspective, a key consideration for SME adoption is the cost and operational feasibility of deploying the IRRS framework. In its tabletop simulation mode, the core instrumentation requirements, including facilitator observation sheets, inject schedules and communication logs, can be implemented with no specialist tooling beyond standard office applications, making the framework accessible to organisations without dedicated security operation functions. A single simulation cycle, including facilitation, scoring and post-exercise review, can typically be completed within a half-day to full-day engagement depending on scenario complexity, comparing favourably to the cost and time investment of external penetration testing or third-party audit engagements. For SMEs with greater maturity, the live emulation mode offers higher fidelity evaluation, and cloud-based Infrastructure as Code templates substantially reduce the infrastructure barrier in this mode. The human-centric telemetry layer adds negligible direct cost, as all three metrics are derived entirely from artefacts already captured by the instrumentation bus. The primary investment

is evaluator time in completing the structured capture sheets during the simulation. These design choices collectively position IRRS as a cost proportionate readiness evaluation mechanism for SMEs, particularly when considered against the documented financial consequences of unpreparedness, including operational disruption, legal liability and the elevated closure rates reported among SMEs following significant cyber incidents (Boswell, 2022).

Overall, the framework's principal contribution lies in its integration of operational realism with scenario-sensitive scoring. The design accounts for the practical constraints faced by SMEs, including limited staffing, incomplete toolsets and decentralised governance, while still delivering an interpretable and repeatable readiness metric. Compared with static policy audits and unstructured tabletop exercises, IRRS offers a more empirical, diagnostic and scalable approach to evaluating IR maturity. Its long-term value will depend on broader application to support reproducible evaluation, scenario extensibility and cross-context benchmarking.

11. Conclusion

This paper introduced and evaluated the IRRS, a structured and scalable framework for empirically assessing organisational IR capability through risk-aware, simulation-based evaluation. By operationalising IR performance across weighted scenarios and observable response actions, IRRS addresses a persistent gap in existing readiness assessments, which often privilege documentation completeness over execution fidelity. The application of IRRS within an SME context demonstrated its effectiveness in exposing operational weaknesses – including delayed containment, informal escalation and underutilisation of available tooling – that are frequently obscured by traditional policy-driven or compliance-oriented approaches.

Beyond readiness measurement, this work extended the IRRS framework through the introduction of human-centric telemetry metrics that provide diagnostic insight into the organisational dynamics underpinning observed performance outcomes. Decision latency, communication entropy and authority drift were shown to function as explanatory lenses that clarify *why* particular IRRS scores emerge under operational pressure, without altering scoring outcomes or maturity classifications. By capturing observable decision-making, coordination efficiency and role adherence behaviours during simulated incidents, these metrics enable deeper interpretation of readiness gaps as systemic properties of the response environment rather than individual shortcomings.

Taken together, IRRS and its associated telemetry shift IR evaluation from static capability claims towards evidence-based performance understanding. This dual-layer approach allows organisations not only to benchmark readiness, but also to identify targeted remediation opportunities aligned with their specific operational constraints and risk profiles. While the current study demonstrates methodological feasibility, further work is required to empirically validate telemetry interpretation bands, assess inter-evaluator reliability and examine metric stability across repeated simulations and organisational contexts. Future research will also explore expanded scenario libraries, longitudinal deployment and integration with SME-scale detection and response tooling. In doing so, IRRS has the potential to evolve into a robust, practitioner-relevant standard for translating IR policy into demonstrable operational resilience.

References

Abid, M. (2025), "IRRS supplementary materials: scenario weighting guide and evaluator rubric", GitHub, available at: https://github.com/tenodex/irrs

Ahmad, A., Maynard, S.B. and Shanks, G. (2015), "A case analysis of information systems and security incident responses", *International Journal of Information Management*, Vol. 35 No. 6, pp. 717-723, doi: 10.1016/j.ijinfomgt.2015.08.001.

Ahmad, A., Maynard, S.B., Desouza, K.C., Kotsias, J., Whitty, M.T. and Baskerville, R.L. (2021), "How can organizations develop situation awareness for incident response: a case study of management practice", *Computers and Security*, Vol. 101, p. 102122, doi: 10.1016/j.cose.2020.102122.

Andy Dodd (2024), "How IASME cyber essentials can protect your SME", Adas-ltd.com, Nov. 13, available at: https://adas-ltd.com/blog/how-iasme-cyber-essentials-can-protect-your-sme/ (accessed 13 May 2025).

Australian Cyber Security Centre (ACSC) (2021), "Small business cyber security guide", Australian Signals Directorate, available at: www.cyber.gov.au/acsc/view-all-content/publications/small-business-cyber-security-guide

Banham, R. (2017), "Cybersecurity threats proliferating for midsize and smaller businesses", *Journal of Accountancy*, Vol. 224 No. 1.

Boswell, R. (2022), "60% Of european SMEs that are cyber-attacked have to close after six months | startups magazine", Startups Magazine, available at: https://startupsmagazine.co.uk/article-60-european-smes-are-cyber-attacked-have-close-after-six-months (accessed 13 March 2025).

Chidukwani, A., Zander, S. and Koutsakis, P. (2024), "Cybersecurity preparedness of small to medium businesses: a Western Australia study with broader implications", *Computers and Security*, Vol. 145, p. 104026, doi: 10.1016/j.cose.2024.104026.

European Union Agency for Cybersecurity (ENISA) (2022), "CSIRT maturity framework – updated and improved", available at: www.enisa.europa.eu/publications/enisa-csirt-maturity-framework

Goings, E., Plesco, R., Nides, D. and Kilman, D. (2016), "10 Common cyber incident response mistakes", KPMG, Apr., available at: https://assets.kpmg.com/content/dam/kpmg/pdf/2016/04/cyber-incident-response.pdf

Ilca, L.F., Lucian, O.P. and Balan, T.C. (2023), "Enhancing cyber resilience for small and medium sized organizations with prescriptive malware analysis, detection and response", *Sensors*, Vol. 23 No. 15, p. 6757, doi: 10.3390/s2315.

International Organization for Standardization (2023), *ISO/IEC 27035-2:2023 – Information Technology – Information Security Incident Management – Part 2: Guidelines to Plan and Prepare for Incident Response*, ISO, Geneva, Switzerland.

Kral, P. (2012), "Incident handler's handbook", SANS Institute, available at: www.sans.org/white-papers/33901/

MITRE Corporation (2025), "CALDERA: automated adversary emulation platform", available at: https://github.com/mitre/caldera

MITRE Engenuity (2025), "MITRE ATT&CK evaluations", available at: https://attackevals.mitre-engenuity.org

National Institute of Standards and Technology (NIST) (2023), "Common vulnerability scoring system v3.1: specification document", available at: www.first.org/cvss/specification-document

National Institute of Standards and Technology (NIST) (2024), "Computer security incident handling guide", NIST Special Publication 800-61 Revision 3, available at: https://csrc.nist.gov/publications

O'Neill, A., Ahmad, A. and Maynard, S. (2021), "Cybersecurity incident response in organisations: a metalevel framework for scenario based training", doi: 10.48550/arXiv.2108.04996.

Oluwawemimo, E. (2024), "The role of artificial intelligence in incident response for digital domain SMEs", Master's thesis, University of Turku.

Open CSIRT Foundation (2023), "SIM3 – security incident management maturity model", available at: https://opencsirt.org/csirt-maturity/sim3-online-tool/

See, P.J., Ong, C., Poon, N., Soh, K.H., Tan, S.G. and Dhaliwal, S. (2024), "Scenario based simulation training for incident management: for whom and how", *Policing: A Journal of Policy and Practice*, Vol. 18, p. paae132, doi: 10.1093/police/paae132.

Verizon (2025), "Data breach investigations report", Verizon Enterprise, May 2025, available at: www.verizon.com/business/resources/reports/dbir/

Corresponding author
Muntathar Abid can be contacted at: montii.abid@student.uts.edu.au