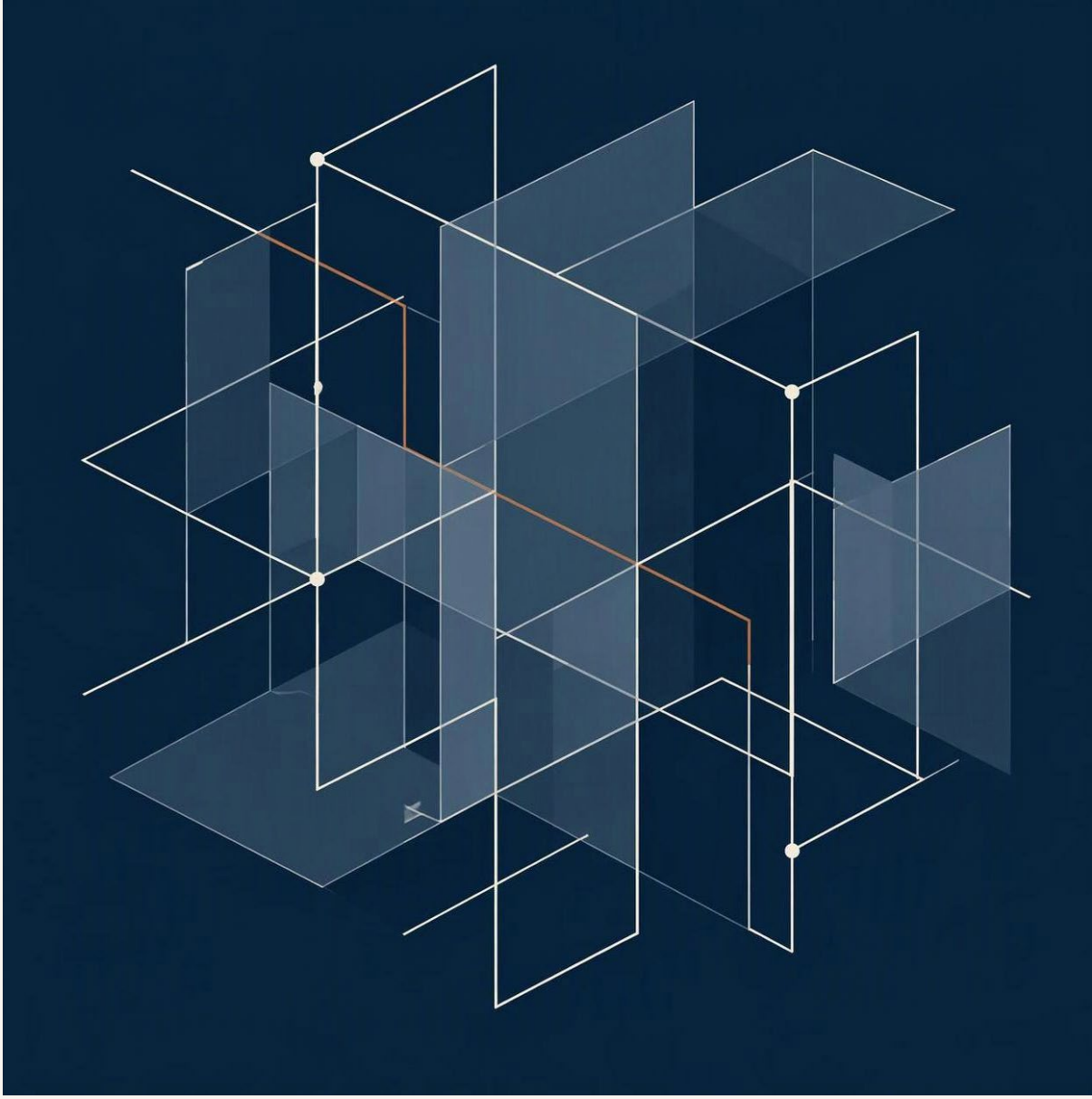
INDEPENDENT
RESEARCH
PUBLICATION



THE INVISIBLE VALUE OF EXECUTION

A New Evidence Framework for Fairer Performance Decisions in Complex Organisations

KARIM GABER SALEH

NOVEMBER 2026

The work that performance reviews struggle to see

"In complex organisations, some of the most valuable contributions are precisely the ones that never become visible as completed tasks."

Performance management becomes difficult when outcomes depend on multiple teams, shared decisions and responsibilities that extend beyond formal reporting lines.

A business program manager may prevent a costly error, resolve a critical dependency or improve a process until recurring problems disappear. These contributions can be consequential, yet difficult to capture through conventional productivity measures.

At the same time, involvement in a successful project does not necessarily demonstrate an individual's contribution. Large budgets, prominent stakeholders and completed milestones can create an appearance of impact that the available evidence does not support.

This publication examines two connected questions:

1. How can organisations assess individual contribution fairly when outcomes are shared and important work is difficult to observe?

2. How can organisations determine whether a manager's performance appraisal is itself fair, consistent and supported by credible evidence?

The proposed framework considers these questions separately while connecting them through a structured review.

EXECUTIVE PERSPECTIVE

Performance has two measurement problems. We usually examine only one.

> *"We evaluate what employees deliver. We rarely examine the integrity of the judgment used to evaluate them."*

The first is measuring contribution.

In complex operational environments, results depend on multiple functions, shared decisions and responsibilities that rarely fit neatly into individual performance objectives.

An employee may influence an important outcome without owning the final decision. Equally, participation in a successful programme does not automatically establish individual contribution.

The second challenge concerns appraisal integrity.

Managers interpret evidence, weigh stakeholder feedback and make consequential judgments about individual performance. These decisions can be affected by incomplete information, inconsistent expectations and differences in visibility.

Subjective judgment remains necessary. The challenge is ensuring that it is disciplined, explainable and open to reasonable scrutiny.

Performance management should examine two separate objects:

01

THE EMPLOYEE'S CONTRIBUTION

What did the individual demonstrably control, influence or improve?

02

THE INTEGRITY OF THE APPRAISAL

Does the manager's conclusion follow from relevant evidence, consistent criteria and a fair review process?

A credible performance system must be able to explain both the contribution and the decision used to assess it.

THE CENTRAL PROPOSITION

Two judgments. Two different standards of evidence.

A credible performance appraisal requires organisations to examine both an employee's demonstrated contribution and the integrity of the decision used to evaluate it.

01

CONTRIBUTION

The employee's demonstrated contribution

CENTRAL QUESTION

What did this individual actually control, influence or improve?

EVIDENCE CONSIDERED

- Agreed responsibilities and accepted outcomes
- Documented interventions and professional judgment
- Collaboration and quality of handoffs
- Verified improvements and preventive actions
- Portfolio complexity and external dependencies

EVALUATION FOCUS

The quality and significance of contribution within the employee's reasonable influence.

02

APPRAISAL INTEGRITY

The quality of the manager's evaluation

CENTRAL QUESTION

Does the appraisal conclusion follow from relevant evidence and consistent standards?

EVIDENCE CONSIDERED

- Previously agreed performance criteria
- Representative and reliable evidence
- Relevant stakeholder observations
- Contradictory evidence and alternative explanations
- Employee response and independent review

EVALUATION FOCUS

The credibility, consistency and fairness of the appraisal decision-making process.

THE PRINCIPLE

A strong contribution does not automatically guarantee a sound appraisal. A procedurally sound appraisal does not automatically establish strong contribution.

Both must be examined.

PROPOSED METHOD This two-object architecture integrates established performance-management concepts into a proposed framework for complex operational roles. Its effectiveness has not yet been independently validated.

PUBLISHED EVIDENCE

How consistent are performance ratings?

Two major meta-analyses published in 2024 provide useful evidence about the reliability of supervisory performance assessments.

FINDING 01 · [1]

r = .65

132 INDEPENDENT SAMPLES

Zhou, Sackett, Shen and Beatty (2024)

An updated meta-analysis of the interrater reliability of supervisory performance ratings.

The study also found meaningful variation by job type:

Managerial positions	r = .57
Nonmanagerial positions	r = .68

The authors caution against relying on one overall reliability estimate across different job categories.

FINDING 02 · [2]

r = .65

22 INDEPENDENT SAMPLES

Speer, Delacruz, Wegmeyer and Perrotta (2024)

Meta-analytical estimates of interrater reliability for direct supervisor performance ratings.

This study applied narrower inclusion criteria focused on direct supervisors. Reported estimates included:

Operational settings	r = .60
Research settings	r = .67

KEY INTERPRETATION

"An interrater reliability estimate of .65 is not equivalent to 65% appraisal accuracy."

These findings concern consistency between raters under studied conditions. They do not establish the percentage of unfair appraisals in any organisation, nor do they demonstrate that the framework proposed in this publication improves appraisal outcomes.

Performance ratings require judgment. The quality of that judgment depends on the evidence, criteria, context and design of the appraisal process.

[1] Zhou et al. (2024). Journal of Applied Psychology, 109(6), 949–970. DOI: 10.1037/apl0001174

[2] Speer et al. (2024). Journal of Applied Psychology, 109(3), 456–467. DOI: 10.1037/apl0001146

RESEARCH INTERPRETATION

A fair outcome and a fair process are different questions.

> *"An employee may challenge not only the rating received, but the way the organisation reached that conclusion."*

Performance appraisals are consequential organisational decisions. They influence career opportunities, recognition, development and sometimes financial rewards. Yet the credibility of an appraisal depends on more than its final rating.

Organisational justice research provides a useful foundation for understanding this distinction. Colquitt and colleagues' influential meta-analysis of organisational justice research distinguishes several dimensions of perceived fairness, including:

DISTRIBUTIVE JUSTICE

The perceived fairness of outcomes, such as ratings, rewards or recognition.

PROCEDURAL JUSTICE

The perceived fairness of the processes used to reach decisions.

INTERPERSONAL JUSTICE

Whether individuals experience respectful and dignified treatment.

INFORMATIONAL JUSTICE

Whether people receive adequate, truthful and appropriate explanations.

These dimensions are related, but they are not interchangeable. A manager may reach a reasonable conclusion while failing to explain the evidence adequately. Conversely, a respectful and transparent process does not automatically make every appraisal conclusion correct.

IMPLICATION FOR PERFORMANCE MANAGEMENT

The proposed framework therefore separates the employee's demonstrated contribution from the integrity of the appraisal process. Appraisal integrity requires organisations to examine whether:

These are proposed operational safeguards informed by organisational justice research. They are not a validated measurement scale.

- Performance expectations were established consistently.
- Relevant evidence received appropriate consideration.
- Materially contradictory information was addressed.
- The employee had a meaningful opportunity to respond.
- The final judgment can be explained and reviewed.

"Fairness requires attention to both the decision and the process that produced it."

[3] Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., and Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. Journal of Applied Psychology.



02

THE INVISIBLE CONTRIBUTION

The most valuable work is not always the most visible.

In complex organisations, outcomes rarely belong to one person.

Some employees create value by resolving dependencies, preventing errors, improving decisions or enabling others to succeed. Their contribution may be substantial even when they do not own the final outcome.

PROPOSED METHOD

Who actually deserves credit?

A successful outcome does not establish individual contribution. Nor does an unsuccessful outcome necessarily indicate poor performance.

Complex operational work often crosses organisational boundaries. Employees may influence important outcomes without having final decision authority.

To make attribution more disciplined, this framework proposes four descriptions of an individual's role.

PROPOSED METHOD | ILLUSTRATIVE EXAMPLES

01	**DIRECTLY OWNED** The employee had agreed authority and responsibility for an identifiable deliverable or decision.	**ILLUSTRATIVE EXAMPLE** *A program manager delivers an agreed governance review within the required timeframe and quality standards.*	**EVIDENCE NEEDED** Agreed responsibility, accepted deliverable, completion record and documented dependencies.
02	**MATERIALLY INFLUENCED** The employee demonstrably improved the quality, timing or options available to a decision-maker without owning the final decision.	**ILLUSTRATIVE EXAMPLE** *A business program manager identifies a material risk, develops alternative options and enables leadership to make an informed decision.*	**EVIDENCE NEEDED** Documented intervention, decision record and traceable evidence of the employee's influence.
03	**SHARED EXECUTION** Multiple individuals or functions contributed to the outcome, making exclusive credit inappropriate.	**ILLUSTRATIVE EXAMPLE** *A complex infrastructure programme reaches a milestone through coordinated contributions from operations, engineering, procurement and finance.*	**EVIDENCE NEEDED** Responsibility records, accepted handoffs and evidence of each participant's material contribution.
04	**EXTERNAL OUTCOME** The final result depended primarily on factors outside the employee's reasonable control.	**ILLUSTRATIVE EXAMPLE** *A project milestone is delayed by an external approval despite timely preparation, escalation and mitigation by the program manager.*	**EVIDENCE NEEDED** Decision dependencies, documented preparation, escalation history and mitigation actions.

"Attribution describes the employee's role in an outcome. It does not automatically determine the employee's performance rating."

> These categories are contextual descriptions, not a hierarchy of employee value. They should not be converted into arbitrary numerical weights or used to excuse failure to fulfil agreed responsibilities. All examples are fictional and demonstrate the intended attribution logic.

RESEARCH INTERPRETATION

The value of problems that never happened

> *"Prevention can create substantial value while leaving remarkably little visible evidence of success."*

Traditional performance measures often recognise completed activities, resolved incidents and delivered milestones. Preventive work creates a different measurement challenge.

When an employee identifies a vulnerability, corrects a control weakness or removes a recurring source of failure, the desired result may be the absence of an incident. But the absence of an incident does not establish that the employee's intervention caused it.

A credible assessment must therefore distinguish observable action from increasingly uncertain claims about avoided consequences.

THREE LEVELS OF PREVENTIVE CONTRIBUTION

01 VERIFIED MITIGATION

What can be established?

The employee identified a specific risk and completed a documented mitigation action.

ILLUSTRATIVE EXAMPLE	EVIDENCE	ASSESSMENT
A program manager identifies a missing operational approval and ensures the control is corrected before implementation.	Risk record, documented intervention, approved corrective action and verification of completion.	**The mitigation action can be recognised as demonstrated contribution.**

02 PLAUSIBLE PREVENTION

What can reasonably be inferred?

Available evidence supports a credible explanation that the intervention reduced the likelihood of an adverse outcome.

ILLUSTRATIVE EXAMPLE	EVIDENCE	ASSESSMENT
The corrected approval process addresses a known failure mechanism that could otherwise have affected implementation.	Documented causal mechanism, relevant operational history and consideration of alternative explanations.	**A plausible contribution to risk reduction may be recognised, with uncertainty clearly stated.**

03 SPECULATIVE AVOIDED LOSS

What cannot be established confidently?

A financial or operational consequence is assigned to an incident that never occurred, without sufficient evidence of its probability or likely severity.

ILLUSTRATIVE EXAMPLE	EVIDENCE	ASSESSMENT
The employee claims that correcting the approval process prevented a hypothetical multimillion-euro loss.	Insufficient evidence to establish the claimed amount or the probability of the hypothetical event.	**Do not present the estimated avoided loss as a realised or verified financial benefit.**

"Recognise the verified intervention. Evaluate the prevention mechanism. Treat hypothetical avoided losses with caution."

PUBLISHED EVIDENCE

Why work across organisational boundaries matters

In complex organisations, important work frequently takes place between formal responsibilities.

Employees coordinate dependencies, obtain information, represent operational requirements and enable decisions involving multiple teams.

Research on team boundary management provides a useful foundation for understanding why these activities matter.

FEATURED RESEARCH

85
PRIMARY STUDIES

10,848
TEAMS

Leicht-Deobald and colleagues (2025). *A Contingency Framework for the Performance Consequences of Team Boundary Management: A Meta-Analysis of 30 Years of Research.*

KEY FINDING

The meta-analysis reported an overall positive association between team boundary management and team performance.

However, the relationship varied according to the nature of the boundary-management activity, its target and who performed it.

RESEARCH INTERPRETATION

Boundary-spanning activities, such as coordination, representation and information gathering, were generally associated with stronger team-performance benefits than boundary-strengthening activities.

The research also found that the effectiveness of boundary management depended on organisational and team context.

IMPLICATION FOR COMPLEX OPERATIONAL ROLES

Business program managers and operational leaders frequently contribute through activities that influence decisions and coordinate work across organisational boundaries. These activities deserve consideration when defining relevant performance criteria.

WHAT THE STUDY EXAMINES

Team-level outcomes.

WHAT IT DOES NOT ESTABLISH

The causal contribution or financial value of any particular employee's intervention.

"Work across organisational boundaries can matter to team performance. Individual credit still requires individual evidence."

[4] Leicht-Deobald, U., et al. (2025). A contingency framework for the performance consequences of team boundary management: A meta-analysis of 30 years of research. *Journal of Management*, 51(2), 704–747. DOI: 10.1177/01492063231206107

When the numbers look objective, but the comparison isn’t.

“Two employees can have different performance percentages without those percentages telling us who performed better.”

FICTIONAL EXAMPLE

95.0% *versus* **90.0%**

Nominal closure rate

Quality-accepted commitment rate

Constructed illustrations of different kinds of work, not research findings or employee data. They cannot establish a valid employee ranking.

Are we measuring performance, or the way work happens to be counted?

Performance indicators depend on choices that are not always visible in the final dashboard.

Who receives the difficult assignments? What counts as completed work? How are reopened cases treated? Were deadlines changed? Did the employee control the outcome?

Before comparing performance, organisations must examine the quality of the underlying measurement and the context in which work was performed.

A precise number can still support an unreliable conclusion.

FICTIONAL WORKED EXAMPLE

The same work. A different result.

A nominal closure rate of 95.0% becomes 89.5% when defect-related reopens are excluded from accepted completions.

THE HIDDEN COST OF COUNTING CLOSURES

FICTIONAL EXAMPLE | PROPOSED MEASUREMENT METHOD

SCENARIO A · HIGH-VOLUME SERVICE OPERATIONS

95.0% → **89.5%**

Nominal closure rate — *Quality-adjusted completion rate*

209 ÷ 220 = 95.0%

220 eligible cases; 209 marked closed on time.

209 − 12 = 197

197 ÷ 220 = 89.5%

However, **12 of those 209 cases were reopened because of defects within their original scope.** Under a consistent quality-acceptance rule, they cannot count as accepted completions.

220 ELIGIBLE CASES



197 quality-accepted + **12 defect-related reopens** + **11 not closed on time** = 220 eligible cases. The 209 nominal closures are the first two segments combined.

SCENARIO B · COMPLEX PROGRAM GOVERNANCE

18 ÷ 20 = 90.0%

20 eligible commitments; 18 quality-accepted commitments delivered on time. Completion rate 90.0%.

COMPARABILITY WARNING

The scenarios involve different work types. The 89.5% and 90.0% figures cannot establish which employee performed better. Work complexity, responsibility, decision authority, external dependencies and assignment opportunities must be considered first.

"Measurement quality matters before performance comparisons begin."

Dashboards may appear objective while concealing choices about:

Which work enters the denominator. Whether closure means acceptance. How defects and reopens are classified. Whether deadlines were legitimately revised. Whether opportunities and responsibilities were comparable.

A credible measurement system must make these rules explicit and auditable.

All values are constructed illustrations of how definitions affect reported results. They are not actual employee records or validated benchmarks.

PROPOSED METHOD

Before comparing performance, compare the work.

A common performance percentage does not make two jobs directly comparable.

Employees may have different responsibilities, case complexity, decision authority, external dependencies and opportunities to demonstrate impact.

A credible appraisal must establish whether the underlying work is sufficiently comparable before drawing conclusions from numerical differences.

THE PORTFOLIO COMPARABILITY GATE

PROPOSED METHOD

01	**WORK ARCHETYPE** *Are the employees performing substantially similar kinds of work?*	Examples: high-volume service delivery versus complex infrastructure governance.
02	**VOLUME AND COMPLEXITY** *How many assignments were received, and how demanding were they?*	Consider ambiguity, technical difficulty and coordination requirements.
03	**RISK AND CONSEQUENCE** *What operational, financial or compliance exposure accompanied the work?*	A large budget describes exposure, not automatically the employee's contribution.
04	**DECISION AUTHORITY** *What could the employee approve or control directly?*	Distinguish formal authority from influence over another decision-maker.
05	**EXTERNAL DEPENDENCIES** *How much did results depend on other teams, suppliers, regulators or external conditions?*	
06	**URGENCY AND WORKLOAD** *Were deadlines, interruptions, unplanned requests and competing priorities materially different?*	
07	**ASSIGNMENT OPPORTUNITY** *Did employees have comparable opportunities to receive visible, high-impact work?*	Consider how managers distributed assignments and whether selection effects could influence reported performance.

DECISION RULE · TWO QUALITATIVE OUTCOMES

Comparable within defined conditions

Relevant work characteristics and measurement rules are sufficiently aligned to support a specific comparison.

Not directly comparable

Material differences or insufficient evidence prevent a defensible numerical ranking.

These are review decisions, not scores: no numerical complexity rating or weighting applies. A comparison may be valid for one measure and inappropriate for another.

"Context changes the question we can fairly ask. It does not automatically add points to an employee's performance."

The gate is a proposed method for improving the interpretation of performance evidence. It has not been independently validated and should be tested for reliability, administrative burden and unintended incentives.

When stakeholder feedback helps, and when it misleads

> *"Stakeholder dissatisfaction is not automatically evidence of poor performance. Stakeholder approval is not automatically evidence of good performance."*

Stakeholder feedback can reveal behaviours that formal indicators miss, such as poor communication, unreliable commitments, disrespectful conduct or failures to collaborate. However, it can also reflect disappointment with a legitimate business decision: a professional who enforces an appropriate control may receive negative feedback from someone whose preferred outcome was rejected.

A sound appraisal must distinguish the quality of the business decision from the quality of the employee's interpersonal conduct.

TWO QUESTIONS, FOUR POSSIBLE OUTCOMES

PROPOSED METHOD

QUALITY OF INTERPERSONAL EXECUTION	WEAK OR UNJUSTIFIED	SOUND AND JUSTIFIED
PROFESSIONAL	**3 WEAK JUDGMENT + PROFESSIONAL CONDUCT** The employee communicates respectfully but makes an inadequately supported decision. *Professional behaviour does not establish decision quality.*	**1 SOUND JUDGMENT + PROFESSIONAL CONDUCT** The employee makes a justified decision and communicates it appropriately. *Positive evidence on both dimensions.*
UNPROFESSIONAL	**4 WEAK JUDGMENT + UNPROFESSIONAL CONDUCT** Both the decision and its execution raise concerns. *Assess each concern using relevant evidence.*	**2 SOUND JUDGMENT + UNPROFESSIONAL CONDUCT** The employee makes a justified decision but behaves disrespectfully or communicates poorly. *Correct judgment does not excuse conduct concerns.*

QUALITY OF BUSINESS OR GOVERNANCE JUDGMENT

FICTIONAL EXAMPLES

SCENARIO A

The unpopular control decision

A program manager refuses an unauthorised exception and explains the decision professionally. The requesting stakeholder expresses dissatisfaction. The appraisal should examine the decision's justification and the employee's conduct independently. Negative sentiment alone does not establish poor performance.

SCENARIO B

The correct decision, poorly handled

A program manager correctly identifies a compliance problem but communicates the decision through insults or intimidation. The underlying judgment may be sound. The conduct concern still requires appropriate consideration.

EVIDENCE STANDARD

Feedback supporting a material appraisal conclusion should identify:

- Relevant first-hand observations
- Specific behaviour and timing
- The expectation involved
- Available context and alternative explanations

A single serious, well-supported concern may matter. Multiple vague endorsements or complaints do not automatically establish a reliable conclusion.

RESEARCH FOUNDATION Research on multisource performance ratings and employee voice supports careful interpretation of feedback. This two-axis framework is a proposed analytical safeguard, not a validated instrument.

[8] Jackson et al. (2020). The implications of unconfounding multisource performance ratings. *Journal of Applied Psychology.* [9] Morrison (2023). Employee voice and silence: Taking stock a decade later. *Annual Review of Organizational Psychology and Organizational Behavior.*

THE COUNTERARGUMENT

When demanding more evidence creates less fairness

> *"An employee who documents everything is not necessarily more valuable than one who quietly does exceptional work."*

An evidence-led appraisal system appears to offer a straightforward improvement: require managers to substantiate their judgments. But this approach carries risks of its own.

Not every important behaviour leaves a reliable documentary trail. Some harmful conduct occurs privately. Some contributions are difficult to capture. Some employees have greater access to influential stakeholders, visible projects or opportunities to present their achievements.

A rigid demand for documentation may therefore reproduce the inequalities that the framework seeks to address.

01

THE VISIBILITY ADVANTAGE

Employees with prominent assignments, influential relationships or greater opportunities for self-promotion may produce more impressive evidence records. A stronger paper trail does not automatically mean stronger performance.

02

THE SILENCING EFFECT

Excessively demanding evidence requirements may discourage legitimate concerns about intimidation, exclusion or other harmful workplace behaviour. Some serious incidents occur without witnesses or contemporaneous documentation. The absence of a perfect record should not automatically invalidate a concern.

03

THE ADMINISTRATIVE BURDEN

If employees must continuously document their contributions, the system may reward administrative effort rather than actual business value. Managers and employees may spend excessive time maintaining evidence repositories. The framework must remain proportionate to the work being assessed.

04

THE POLITICS OF REVIEW

Independent review does not automatically eliminate institutional bias. Reviewers may share the same incentives, assumptions or organisational relationships as the original decision-maker. A formally documented appraisal may still contain selective evidence or weak reasoning.

THE REQUIRED RESPONSE

A credible evidence-led framework must therefore:

- Accept appropriate qualitative evidence.
- Protect confidential reporting and legitimate complaints.
- Examine contradictory evidence and alternative explanations.
- Consider differences in employees' opportunities to demonstrate impact.
- Limit documentation requirements to proportionate, decision-relevant evidence.
- Provide independent review and meaningful opportunities to challenge material conclusions.

"The purpose of stronger evidence is better judgment, not more paperwork."

RESEARCH POSITION This page identifies material risks in the proposed framework. These safeguards are design proposals; they have not yet been shown through independent testing to eliminate unfairness or improve appraisal outcomes.

The Contribution and Appraisal Integrity Architecture

Four distinct layers of evidence. Two separate objects of evaluation. No composite employee score.

OBJECT OF EVALUATION 1 · EMPLOYEE CONTRIBUTION AND ITS CONTEXT

A

INDIVIDUAL CONTRIBUTION

What did the employee demonstrably control, influence or improve?

Assess the individual's demonstrated contribution within agreed responsibilities and reasonable influence.

EVIDENCE DOMAINS

- Execution reliability
- Professional judgment and intervention
- Collaboration and handoff quality
- Structural and preventive improvements

B

TEAM EFFECTIVENESS

How effectively did the wider operation perform?

Understand collective performance without automatically assigning shared outcomes to individual employees.

EVIDENCE DOMAINS

- Dependency resolution
- Decision readiness
- Quality-accepted outcomes
- Rework and control exposure
- Sustained operational improvement

C

PORTFOLIO AND OPPORTUNITY CONTEXT

Under what conditions did the employee work?

Establish the context necessary to interpret contribution and determine whether comparisons are defensible.

EVIDENCE DOMAINS

- Work type and complexity
- Risk and consequence
- Decision authority
- External dependencies
- Assignment and visibility opportunities

OBJECT OF EVALUATION 2 · THE APPRAISAL PROCESS

D

APPRAISAL INTEGRITY

Does the manager's conclusion follow from credible evidence and fair procedures?

Evaluate the integrity of the appraisal decision itself.

EVIDENCE DOMAINS

- Agreed criteria
- Evidence quality and representativeness
- Contradictory information
- Employee response
- Calibration and independent review

"The employee's contribution and the manager's appraisal are separate subjects of evaluation. Both require evidence."

SUPPORTING NOTE This architecture integrates established concepts from performance measurement, organisational justice and operational governance. Its application to complex business program management is proposed and has not yet been independently validated.

Individual contribution and collective performance

> *"An employee can contribute substantially to a team outcome without owning it. Equally, strong team performance does not automatically establish strong individual performance."*

A

LAYER A · INDIVIDUAL CONTRIBUTION — PROPOSED METHOD

What did the employee demonstrably control, influence or improve?

01 EXECUTION RELIABILITY

Assess eligible commitments accepted to agreed quality and timing standards. Account for documented dependencies, legitimate deadline changes and actual decision authority.

02 PROFESSIONAL JUDGMENT AND INTERVENTION

Examine how the employee diagnosed problems, framed alternatives, escalated proportionately and followed through on decisions. The number of escalations alone does not establish good judgment.

03 COLLABORATION AND HANDOFF QUALITY

Evaluate the clarity, reliability and usefulness of contributions across organisational boundaries. Consider accepted handoffs, avoidable rework and relevant behaviour-specific stakeholder feedback.

04 STRUCTURAL IMPROVEMENT

Recognise verified changes that reduce recurring work, improve controls or enable sustainable operational improvements. Distinguish demonstrated improvement from speculative avoided losses.

B

LAYER B · TEAM EFFECTIVENESS — PROPOSED METHOD

How effectively did the wider operation perform?

OPERATIONAL INDICATORS

- Ageing and resolution of material dependencies.
- Time required to reach decision readiness.
- Quality-accepted service outcomes.
- Preventable rework.
- Unresolved control exposure.
- Sustained adoption of improvements.

IMPORTANT DISTINCTION

Team indicators describe collective operational performance. They provide context for individual assessment but must not automatically determine an employee's appraisal.

A team may perform well despite a particular individual's weak contribution. A capable employee may also contribute effectively within a team affected by external constraints or wider operational failures.

"Individual contribution requires attribution. Team performance requires an understanding of the wider system."

NOTE The evidence domains are illustrative components of a proposed assessment architecture. They are not empirically validated KPIs or universal measures of employee performance, and team indicators are not converted into individual scores.

Context explains the work. Integrity tests the judgment.

> *"Fair performance assessment requires an understanding of the employee's operating conditions and a credible explanation of the manager's conclusions."*

C

LAYER C · PORTFOLIO AND OPPORTUNITY CONTEXT

PROPOSED METHOD

Under what conditions did the employee perform?

WORK CHARACTERISTICS
Type, volume, complexity and ambiguity of assigned responsibilities.

RISK AND CONSEQUENCE
Operational, financial, commercial or compliance exposure associated with the work.

DECISION AUTHORITY
The distinction between decisions controlled directly and outcomes influenced indirectly.

EXTERNAL DEPENDENCIES
Reliance on other teams, suppliers, regulators or external decision-makers.

ASSIGNMENT OPPORTUNITY
Differences in access to visible, strategically important or high-impact assignments.

ANALYTICAL PURPOSE

Portfolio context determines which comparisons are defensible. It must not become an arbitrary numerical multiplier or an automatic reward for prestigious projects.

D

LAYER D · APPRAISAL INTEGRITY

PROPOSED METHOD

Does the manager's evaluation follow from credible evidence, consistent expectations and a fair decision-making process?

01 **PRE-AGREED CRITERIA**
Were expectations established clearly, consistently and before the appraisal period?

02 **EVIDENCE QUALITY**
Does the conclusion rely on relevant, representative and reasonably reliable evidence?

03 **CONTRADICTORY INFORMATION**
Did the manager consider material evidence that challenges the proposed conclusion?

04 **EMPLOYEE RESPONSE**
Did the employee have an appropriate opportunity to understand and respond to material concerns?

05 **CALIBRATION AND REVIEW**
Can a reviewer understand the reasoning and examine its consistency with relevant standards?

ANALYTICAL PURPOSE **Appraisal integrity assesses the quality of the evaluation process. It does not measure the manager's personality or automatically establish whether the underlying employee rating is correct.**

"Managers should be accountable for explaining consequential performance judgments, including how they considered evidence that challenged their conclusions."

NOTE A review process cannot automatically eliminate organisational politics or poor judgment; its effectiveness depends on appropriate independence, credible escalation routes and organisational safeguards. This architecture is a conceptual research proposal, not a validated appraisal instrument. Any actual review or remedy must operate within applicable employment policies, contractual rights and local law.

PROPOSED METHOD

Making the manager's judgment reviewable

> *"An evidence-led appraisal requires clear expectations, representative evidence and a decision trail that can withstand reasonable scrutiny."*

Performance evaluations frequently rely on managerial interpretation.

The proposed Appraisal Integrity Protocol introduces seven stages designed to make consequential performance judgments more transparent, consistent and open to appropriate review. This page presents the first four stages.

THE SEVEN-STAGE APPRAISAL INTEGRITY PROTOCOL — PART ONE: ESTABLISHING THE EVIDENCE

01 EXPECTATIONS

Establish the criteria against which contribution will be assessed.

REQUIRED PRACTICE At the beginning of the review period, the manager and employee document role outcomes, major responsibilities, performance standards and foreseeable dependencies. Material changes should be dated and acknowledged rather than introduced retrospectively.

EVIDENCE ARTIFACT
Agreed performance expectations and documented changes.

02 PORTFOLIO CONTEXT

Document the operating conditions that affect performance interpretation.

REQUIRED PRACTICE Record work assignments, case mix, complexity, decision authority, external dependencies and material differences in employee opportunities. Assignment changes should remain traceable.

EVIDENCE ARTIFACT
Portfolio context and assignment record.

03 REPRESENTATIVE EVIDENCE

Evaluate contribution using a balanced evidence sample.

REQUIRED PRACTICE Review relevant operational records and representative cases covering successful delivery, setbacks, routine work, prevention, improvements and contributions from other functions. Avoid relying exclusively on employee-selected achievements or manager-selected failures.

EVIDENCE ARTIFACT
Traceable performance evidence record.

04 BEHAVIOUR-SPECIFIC FEEDBACK

Include relevant stakeholder observations without treating popularity as performance.

REQUIRED PRACTICE Seek appropriately diverse first-hand perspectives. Record the behaviour observed, the relevant period, applicable expectations and material context. Protect confidentiality where necessary.

EVIDENCE ARTIFACT
Qualified stakeholder feedback record.

"Before a manager reaches a consequential conclusion, the evidence should be relevant, representative and interpreted in context."

CONTINUES Part Two, page 20: contradiction reconciliation, calibration, independent review and remedies. This is a research-informed governance design, not a validated organisational appraisal process; it has not been implemented or shown to improve fairness.

PROPOSED METHOD

A judgment should withstand scrutiny

The Appraisal Integrity Protocol, Part Two

> *"An appraisal becomes more accountable when material conclusions can be explained, challenged and reviewed."*

FROM EVIDENCE TO AN ACCOUNTABLE DECISION — PART TWO: STAGES 05–07

05

CONTRADICTION RECONCILIATION

Ensure that disputed or materially consequential judgments receive appropriate scrutiny.

REQUIRED PRACTICE For a material contested conclusion, the manager should examine: the evidence supporting the conclusion; relevant contradictory information; context and alternative explanations; and the distinction between business judgment and interpersonal conduct. The employee should have a meaningful opportunity to respond to the substance of material concerns under applicable procedures.

EVIDENCE ARTIFACT

Documented reasoning, contradictory evidence considered and employee response record.

06

CALIBRATION AND DECISION

Assess consistency and reach an explainable performance conclusion.

REQUIRED PRACTICE An appropriately trained and authorised reviewer or calibration panel examines whether the manager applied relevant standards consistently. Calibration should consider comparable responsibilities, evidence quality and portfolio context. It should not simply force employees into predetermined rating distributions. The final decision should explain: the material criteria applied; the evidence supporting the judgment; how contradictory information was considered; and any relevant uncertainty.

EVIDENCE ARTIFACT

Documented appraisal rationale and calibration record.

07

AUDIT AND REMEDY

Test whether the appraisal process operated with sufficient integrity.

REQUIRED PRACTICE An authorised HR or independent process owner reviews a sample of appraisal decisions, examining: evidence coverage and representativeness; unexplained deviations from agreed criteria; material assignment and opportunity differences; handling of contradictory information; and employee response and appeal processes. Where the organisation's applicable rules permit, deficient decisions may be returned for correction or escalated through an appropriate independent process.

EVIDENCE ARTIFACT

Appraisal integrity review record and documented corrective actions.

POSSIBLE REVIEW PATHWAY

MANAGER'S CONCLUSION	EVIDENCE AND REASONING REVIEW	EMPLOYEE RESPONSE	AUTHORISED REVIEW	DECISION CONFIRMED, REVISED OR ESCALATED

Possible review stages and outcomes only; not guaranteed rights or universal procedures.

"Managerial discretion carries a responsibility to explain how consequential judgments were reached."

NOTE A documented process cannot automatically eliminate bias, organisational politics or selective reasoning; its effectiveness depends on credible review authority, appropriate independence and safeguards against retaliation. Any actual remedy must operate within applicable organisational policies, employment contracts and local law. This is a research-informed governance proposal that has not been independently validated or shown to improve appraisal outcomes; review and appeal rights vary between organisations.

AI adoption is not the same as productivity

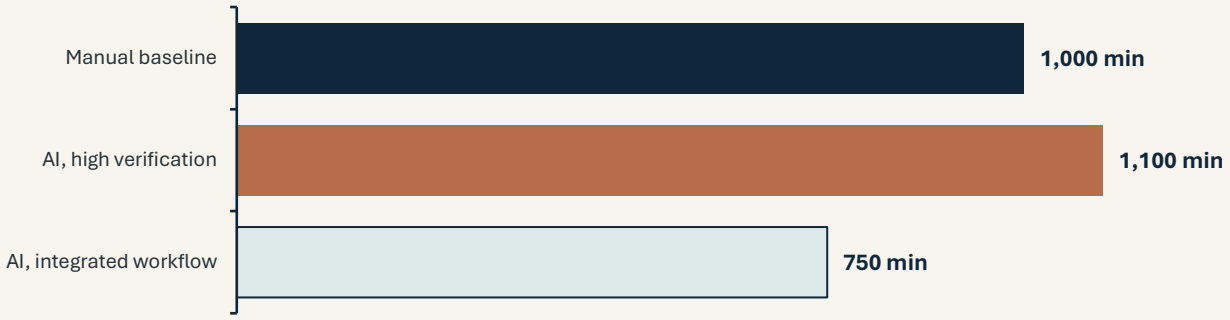
> *"Using AI more often does not establish that an employee creates more value."*

AI can accelerate drafting, information retrieval, reporting and operational analysis. However, the time saved during one activity may be offset by verification, correction, rework or additional coordination.

Counting AI prompts, generated documents or automated tasks therefore provides an incomplete picture of productivity. A credible assessment must examine the complete workflow and the quality of accepted results.

WHEN FASTER DRAFTING DOES NOT SAVE TIME

FICTIONAL EXAMPLE



SCENARIO 01 · MANUAL BASELINE

1,000 min

TOTAL EFFORT · 50 × 20 minutes

Reference point for comparison.

SCENARIO 02 · AI WITH HIGH VERIFICATION EFFORT

1,100 min

TOTAL EFFORT · 50 × 22 minutes

+100 minutes vs baseline: a 10% increase in total effort.

SCENARIO 03 · AI WITH EFFECTIVE WORKFLOW INTEGRATION

750 min

TOTAL EFFORT · 50 × 15 minutes

−250 minutes vs baseline: a 25% reduction in total effort.

INTERPRETATION

Evaluate AI productivity across the complete workflow.

- Drafting and preparation time.
- Verification and correction.
- Accepted output quality.
- Maintenance and training effort.
- Work transferred to other colleagues.
- Rework and operational risk.

"Measure the work that remains after AI has finished its part."

NOTE Fictional arithmetic examples constructed for illustration, not empirical findings; the 10% increase and 25% reduction do not reflect measured AI productivity in real organisations. They assume comparable tasks and acceptable output quality. The comparison alone does not establish realised financial savings, and time saved does not automatically become financial value.

When does AI productivity become business value?

> *"Time saved during a task is only the beginning of an evidence-based productivity claim."*

AI productivity benefits can be overstated when organisations treat improvements at one stage of a workflow as proof of wider business value. A reduction in drafting time may disappear once verification and rework are included.

Even genuine reductions in total workflow effort do not automatically create additional capacity or financial savings. This framework distinguishes four progressively more demanding claims.

FOUR LEVELS OF AI BENEFIT EVIDENCE

PROPOSED METHOD · NOT A VALIDATED MODEL

01	**STEP-LEVEL TIME SAVINGS** *AI reduces the time required for a particular activity.*	**EVIDENCE REQUIRED** Comparable before-and-after task measurements.	**LIMITATION** **Time saved at one step may be offset elsewhere.**
02	**NET WORKFLOW TIME RELEASED** *The complete workflow requires less time while maintaining acceptable quality.*	**EVIDENCE REQUIRED** End-to-end time measurement, including verification, corrections and rework.	**LIMITATION** **Reduced effort does not automatically establish productive redeployment.**
03	**CAPACITY DEMONSTRABLY REDEPLOYED** *The organisation uses released capacity for additional valuable work.*	**EVIDENCE REQUIRED** Documented changes in workload, throughput or completed priorities.	**LIMITATION** **Additional capacity does not automatically translate into realised financial benefit.**
04	**REALISED FINANCIAL BENEFIT** *The organisation achieves a demonstrable financial outcome attributable to the intervention.*	**EVIDENCE REQUIRED** Credible financial records, relevant costs, an established baseline and consideration of alternative explanations.	**LIMITATION** **Savings or benefits must be supported by evidence rather than hypothetical extrapolation.**

"Each higher level requires additional evidence. None follows automatically from the previous one."

AI IN EMPLOYEE APPRAISAL: NECESSARY SAFEGUARDS

PROPOSED SAFEGUARDS

- Do not rank employees by AI prompt volume or tool usage.
- Do not treat AI-generated output as automatically accepted work.
- Do not infer employee sentiment or performance from personal AI interactions.
- Maintain appropriate human accountability for consequential appraisal decisions.
- Apply data minimisation, confidentiality and appropriate access controls.
- Obtain authorised HR, privacy and legal review before deploying workplace evaluation systems.

REGULATORY CONTEXT The EU AI Act identifies certain AI systems used to monitor or evaluate workers within its employment-related high-risk categories. Classification requires consideration of Annex III together with Article 6 and the applicable legal provisions. Not every AI tool used by an employer is automatically high-risk, and these safeguards do not guarantee legal compliance.

SOURCES [11] European Union Artificial Intelligence Act, Annex III, section 4(b), read with Article 6. [12] OECD (2025). How Widespread Is Algorithmic Management in Workplaces?

RESEARCH VALIDATION

What would it take to prove this framework works?

> *"A credible pilot must be capable of disproving our preferred conclusions."*

The proposed framework has undergone internal conceptual development and fictional scenario testing. Those exercises help identify design weaknesses but do not establish that the framework improves appraisal fairness, reliability or organisational outcomes.

The next research stage requires independent evaluation using appropriately authorised and protected evidence.

SIX TESTS FOR A CREDIBLE PILOT

PROPOSED PILOT METHOD

01 SCOPE AND SAMPLING
Include at least two different operational work archetypes. Define eligibility, observation periods and case-selection rules before testing.

02 DATA INTEGRITY
Independently reconstruct a sample of assignments, deadlines, accepted results, reopens and material interventions. Determine whether proposed measures can be reproduced consistently.

03 REVIEWER CONSISTENCY
Use at least two independent human reviewers to examine the same anonymised cases. Compare their conclusions, document disagreements and examine whether anchored criteria improve consistency.

04 FAIRNESS AND ATTRIBUTION
Test whether irrelevant factors such as portfolio prestige, stakeholder sentiment or assignment visibility affect judgments. Investigate whether contribution is attributed appropriately in shared outcomes.

05 EMPLOYEE SAFEGUARDS
Evaluate whether employees can understand and contest material findings while protecting legitimate confidentiality. Include cases involving credible criticism as well as potentially unjustified negative feedback.

06 OPERATIONAL FEASIBILITY
Measure the administrative effort required from employees, managers and reviewers. Assess whether the framework is practical enough for sustained organisational use.

WHAT WOULD CAUSE US TO REVISE OR REJECT IT?

- Key performance denominators cannot be independently reconstructed.
- Measurement rules systematically reward easier case selection.
- Independent reviewers continue to disagree materially after appropriate training.
- The process disadvantages employees with fewer opportunities for visible work.
- Legitimate complaints become harder to raise or assess.
- Administrative burden outweighs demonstrated usefulness.

RETAIN

Evidence supports further testing or limited use within defined conditions.

REVISE

The framework shows promise but requires changes to measures, safeguards or procedures.

REJECT

Material weaknesses make the proposed approach unsuitable for the intended use.

"The framework earns credibility through independent testing, not through the elegance of its design."

METHODOLOGICAL DISCLOSURE The framework remains a research-informed conceptual proposal. Its initial scenarios and calculations were constructed for methodological testing. No independent field study, real-world validation or demonstrated fairness improvement is claimed, and no pilot has been conducted.

What should organisations do differently?

> *"Performance management should examine both what employees contribute and how managers reach their conclusions."*

Complex operational work challenges conventional performance measurement. Individual contribution is frequently distributed across teams, influenced by external decisions and expressed through activities that are difficult to observe. Yet making performance assessment more quantitative does not automatically make it more accurate or fair.

The proposed framework offers a different approach: distinguish individual contribution from collective outcomes, examine operating context and make consequential appraisal judgments more reviewable.

EIGHT ACTIONS WORTH TESTING

PROPOSED PRACTICES

01 DEFINE EXPECTATIONS EARLY
Establish clear, role-relevant performance criteria before the review period and document material changes.

02 SEPARATE CONTRIBUTION FROM OUTCOMES
Distinguish what employees directly controlled, materially influenced or contributed to collectively.

03 MEASURE ACCEPTED QUALITY
Assess completed work against appropriate acceptance standards rather than relying exclusively on activity or closure volume.

04 EXAMINE PORTFOLIO CONTEXT
Consider complexity, dependencies, decision authority and assignment opportunities before comparing employees.

05 IMPROVE STAKEHOLDER FEEDBACK
Seek relevant first-hand observations and behaviour-specific examples rather than personality judgments or popularity measures.

06 REQUIRE EXPLAINABLE APPRAISALS
Managers should substantiate material conclusions using relevant criteria, representative evidence and documented reasoning.

07 PROVIDE CREDIBLE REVIEW
Establish proportionate opportunities to address contradictory evidence, challenge material conclusions and obtain authorised independent review.

08 ASSESS AI THROUGH REAL IMPACT
Evaluate AI-assisted work using accepted quality, net workflow effort and demonstrable business outcomes rather than tool usage.

"The objective is not to replace managerial judgment with formulas. It is to make that judgment more disciplined, transparent and accountable."

The Contribution and Appraisal Integrity Architecture remains a research-informed proposal. Its practical value must be established through independent testing across different operational environments. The next step is empirical evaluation, not an assumption that the framework has already solved the problem.

Make contribution visible.
Make judgment accountable.

Methodology, limitations and disclosure

> *"This publication proposes a research-informed framework. It does not claim to present a validated performance-appraisal instrument."*

01 RESEARCH APPROACH

This publication uses a targeted narrative synthesis and conceptual framework-development approach.

Relevant published research was examined across supervisory performance ratings, organisational justice, stakeholder feedback, employee voice, team boundary management, evaluation methodology, multidimensional productivity and algorithmic management.

It draws on peer-reviewed academic publications and authoritative institutional guidance. It is not a systematic literature review or an exhaustive assessment of all available research.

02 FRAMEWORK DEVELOPMENT

The Contribution and Appraisal Integrity Architecture was developed by integrating established concepts into a proposed approach for complex operational and business program management roles.

The framework distinguishes individual contribution, team effectiveness, portfolio context and appraisal integrity, and combines them with a proposed seven-stage appraisal review process.

The architecture and protocol are original integrations of existing concepts. Their effectiveness has not been independently established.

03 SYNTHETIC TESTING

The development process included fictional scenarios designed to challenge the framework. These explored shared outcomes, uneven work complexity, stakeholder feedback, preventive contributions, quality-adjusted completion measures and AI productivity claims.

The scenarios were constructed by the research author. They were not derived from employee records or independently collected field data.

The exercises show how the proposed assessment rules behave under selected assumptions. They do not establish real-world reliability, validity, fairness or organisational effectiveness.

04 LIMITATIONS

EVIDENCE TRANSFER Some supporting studies examine adjacent research domains rather than business program management specifically.

ATTRIBUTION UNCERTAINTY Individual contribution can remain difficult to distinguish from collective outcomes.

SUBJECTIVE JUDGMENT Evidence requirements cannot eliminate interpretation, institutional incentives or potential bias.

DATA AVAILABILITY Operational records may be incomplete, inconsistent or unsuitable for reconstructing contribution.

VALIDATION No independent empirical pilot, formal reviewer-reliability study or prospective organisational evaluation has been completed.

GENERALISABILITY The proposed framework has not been validated across industries, employers, jurisdictions or job categories.

05 RESEARCH INDEPENDENCE AND DISCLOSURE

This is independent research authored by Karim Gaber Saleh. Professional experience in complex operational environments informed the research questions.

No Microsoft personnel records, grievance materials, confidential operational data or commercial records were used as an empirical dataset. The publication does not represent the position, policies or endorsement of any current or former employer.

All worked employee scenarios, numerical performance comparisons and AI productivity calculations are fictional unless explicitly identified as published external research.

06 FUTURE RESEARCH

The next stage should involve independent reviewers, appropriately authorised evidence and a prospective pilot examining measurement reliability, appraisal consistency, fairness safeguards and administrative burden.

The framework should be revised or rejected where evidence fails to support its intended use.

Sources and further reading

Peer-reviewed research and authoritative institutional publications informing the proposed framework.

01

Zhou, Y., Sackett, P. R., Shen, W., & Beatty, A. S. (2024). **An updated meta-analysis of the interrater reliability of supervisory performance ratings.** *Journal of Applied Psychology*, 109(6), 949–970.

https://doi.org/10.1037/apl0001174

02

Speer, A. B., Delacruz, A. Y., Wegmeyer, L. J., & Perrotta, J. (2024). **Meta-analytical estimates of interrater reliability for direct supervisor performance ratings: Optimism under optimal measurement designs.** *Journal of Applied Psychology*, 109(3), 456–467.

https://doi.org/10.1037/apl0001146

03

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). **Justice at the millennium: A meta-analytic review of 25 years of organizational justice research.** *Journal of Applied Psychology*.

https://pubmed.ncbi.nlm.nih.gov/11419803/

04

Leicht-Deobald, U., Backmann, J., de Vries, T. A., Weiss, M., Hohmann, S., Walter, F., van der Vegt, G. S., & Hoegl, M. (2025). **A contingency framework for the performance consequences of team boundary management: A meta-analysis of 30 years of research.** *Journal of Management*, 51(2), 704–747.

https://doi.org/10.1177/01492063231206107

05

UK Government. **Magenta Book Annex A: Analytical Methods for Use Within an Evaluation.** Contribution analysis.

https://www.gov.uk/government/publications/the-magenta-book/magenta-book-annex-a-analytical-methods-for-use-within-an-evaluation-html

06

Chartered Institute of Personnel and Development (CIPD). **Performance Management.** Evidence-informed factsheet.

https://www.cipd.org/en/knowledge/factsheets/performance-factsheet/

Sources and further reading

References 07–12

07 Chartered Institute of Personnel and Development (CIPD). (2022). **Performance Feedback: An Evidence Review.**
https://www.cipd.org/uk/knowledge/evidence-reviews/performance-feedback

08 Jackson, D. J., Michaelides, G., Dewberry, C., Schwencke, B., & Toms, S. (2020). **The implications of unconfounding multisource performance ratings.** *Journal of Applied Psychology*, 105(3), 312–329.
https://doi.org/10.1037/apl0000434

09 Morrison, E. W. (2023). **Employee voice and silence: Taking stock a decade later.** *Annual Review of Organizational Psychology and Organizational Behavior*, 10, 79–107.
https://doi.org/10.1146/annurev-orgpsych-120920-054654

10 Forsgren, N., Storey, M.-A., Maddila, C., Zimmermann, T., Houck, B., & Butler, J. (2021). **The SPACE of Developer Productivity: There's more to it than you think.** *ACM Queue*, 19(1).
https://www.microsoft.com/en-us/research/publication/the-space-of-developer-productivity-theres-more-to-it-than-you-think/

11 European Commission. **AI Act Service Desk.** Annex III, Section 4(b), read together with Article 6 and the applicable implementation provisions.
https://ai-act-service-desk.ec.europa.eu/en/ai-act/annex-3

12 Organisation for Economic Co-operation and Development (OECD). (2025). **How Widespread Is Algorithmic Management in Workplaces?** *OECD Publishing.*
https://doi.org/10.1787/cda7a114-en

ABOUT THIS PUBLICATION

The Invisible Value of Execution is an independent research publication by Karim Gaber Saleh.

The proposed Contribution and Appraisal Integrity Architecture is a research-informed conceptual framework. Its effectiveness has not yet been independently validated.

November 2026.