2026 BENCHMARK REPORT

# The Confidence Gap

HR leaders trust their performance ratings.
Most can't prove them.

**Based on 230 People & HR Leaders · Tech Companies of 50–200**

**66%**
TRUST RATINGS ARE ACCURATE

**27%**
COULD DEFEND EVERY RATING

**65%**
HAVE SEEN AN UNSUPPORTED SCORE

WHAT'S INSIDE

# Table of Contents

FOREWORD

# The Gap Between Trust and Proof

Ask an HR leader whether their ratings are fair, and the answer is quick: yes. Ask whether a manager could prove any single rating with evidence, and the room goes quiet.

That silence is this report. Two thirds of the 230 People leaders we surveyed say ratings are trusted. Only one in four say a manager could defend every score with goal evidence. Most have personally watched a rating the evidence didn't support — many more than once.

It isn't that performance management is broken. It's that confidence in the system runs well ahead of the evidence behind any one score — because the goal record that would settle a challenge lives somewhere else, on another calendar, rarely in the room when the rating gets written.

**66%**
Say ratings are trusted

**27%**
Say every rating could be defended

**65%**
Have seen an unsupported score

This report measures that gap — and shows what the organizations that have closed it do differently.

EXECUTIVE SUMMARY

# The Headline Findings

**66%**

of HR leaders say ratings are highly trusted — yet only 27% say nearly all managers could defend every rating with goal evidence.

**65%**

Have seen a rating the goal evidence didn't support

**92%**

Don't run performance inside their goal system

01 **Ratings are trusted in aggregate and thin in particular.**

66% report high trust, but only 27% say managers could point to documented goal evidence if a score were challenged.

02 **Most leaders have watched it happen.**

65% have seen a rating the goal evidence didn't support — 39% more than once. Not a hypothetical; a recurring event.

03 **Even where goals should drive the score, something else creeps in.**

60% say goal delivery should most influence a rating; only 55% say it actually does. Recency absorbs the difference.

04 **The goals and the review run on separate calendars.**

61% run performance separately from — or disconnected from — their goal cycle.

05 **AI is entering reviews faster than the record it needs.**

54% already use AI in reviews, but only 8% run performance on the goal tool that would give it a real record.

SECTION 1

# The Confidence Paradox

**27%**
Could defend every rating with evidence

Trusted system, undefendable scores

The clearest signal in this study isn't a single number — it's the distance between two. Leaders are confident their ratings are fair. They are far less confident any given rating could be defended: “about half” was the most common answer when asked how many managers could produce evidence for a score — not “nearly all.” The same people hold both views at once.

## Trust is high. Defensibility is not.

| | |
|---|---|
| Employees highly trust ratings | 66% |
| Nearly all managers could defend a rating | 27% |

> *The paradox lives inside a single respondent. Even among the 66% who say ratings are highly trusted, 36% admit only a minority of managers could defend a rating — and 62% have personally seen one the evidence didn't support. The same person believes the system and doubts the score.*

And this isn't theoretical. Asked whether they'd personally seen a rating the goal evidence didn't support, two thirds said yes:

**39%**
Have seen it more than once

**27%**
Have seen it once

**34%**
Say never — or lack the record to notice

The uncomfortable reading: you cannot catch a rating the evidence doesn't support if there is no evidence to check it against.

SECTION 2

# What the Score Is Built On

**45%**
Say something other than goals drives the rating

The aggregate holds. The individual doesn't.

Leaders know what a rating should rest on. Asked what **should** most influence a score, 60% named delivery against goals. Asked what **actually** does, only 55% said the same. The gap is small — but the shape underneath it is the story.

| WHAT MOST INFLUENCES A RATING | SHOULD | ACTUALLY DOES |
|---|---|---|
| Delivery against goals, full period | **60%** | **55%** |
| What they delivered most recently | 20% | 28% |
| How visible / present they are | 15% | 13% |
| How well they advocate for themselves | 6% | 5% |

The leak is recency. “Most recent work” climbs from a hoped-for 20% to an admitted 28% — the single largest should-versus-does movement in the data. Nearly half of leaders concede something other than full-period goal delivery is what actually moves the number.

> *Goals hold up in aggregate — across a whole organization, delivery is still the leading input. The failure is per-rating: on any individual score, the odds that recency or presence quietly displaced the goal record are close to even.*

The mechanism is memory. Only 52% of leaders say a manager's recall covers the full period evenly; for the rest, the review is written from the last month or two — exactly where recency takes over from record.

SECTION 3

# The Bias That Fills the Gap

**68%**

Have seen early work forgotten by review time

What rushes in where the record is missing

Where a complete goal record is absent, bias fills the vacuum. Two patterns show up clearly: work done early gets forgotten, and work done out of sight gets discounted. Both are symptoms of the same missing record.

## The vanished quarter

More than two thirds have watched a strong performer rated lower because their best work happened early and faded by review time.


## Out of sight, marked down

**62%**
Say remote / hybrid staff are at least somewhat disadvantaged

**26%**
Say meaningfully disadvantaged — out of sight lowers scores

**33%**
Say it makes no difference

> *Recency and proximity bias are usually treated as human failings to train away. The data suggests they're structural: both recede when a complete, time-stamped goal record replaces memory as the basis for the score. You don't coach away a vanished quarter — you record it.*

SECTION 4

# Two Calendars

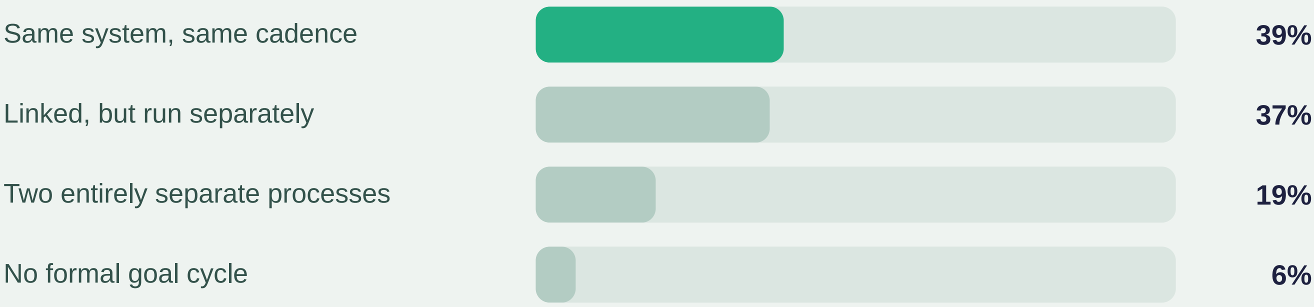
Why the goals aren't in the room

**61%**

Run goals & reviews separately or disconnected

The confidence gap has a structural cause. For most organizations, the goal cycle and the review cycle are two different processes, in two different places, on two different calendars. When the review happens, the goals have to be fetched — often reconstructed — rather than simply read.

## How connected are the two cycles?




> *The proof is in the split. Teams running goals and reviews as one integrated cycle are more than twice as likely to say nearly all managers could defend a rating — 42% versus 17% — and far less likely to have seen an unsupported one (53% versus 74%). Integration isn't a nicety. It's what makes a rating provable.*

It isn't that leaders don't value goals in ratings — 87% use goal achievement as a direct determinant or one factor among several. The problem is access, not intent: the input everyone agrees matters is the one kept hardest to reach at the moment of the rating. 35% say they have to reconstruct the person's goals at review time rather than read them from one place — and reconstruction is where the record becomes memory again.

SECTION 5

# AI Without a Record

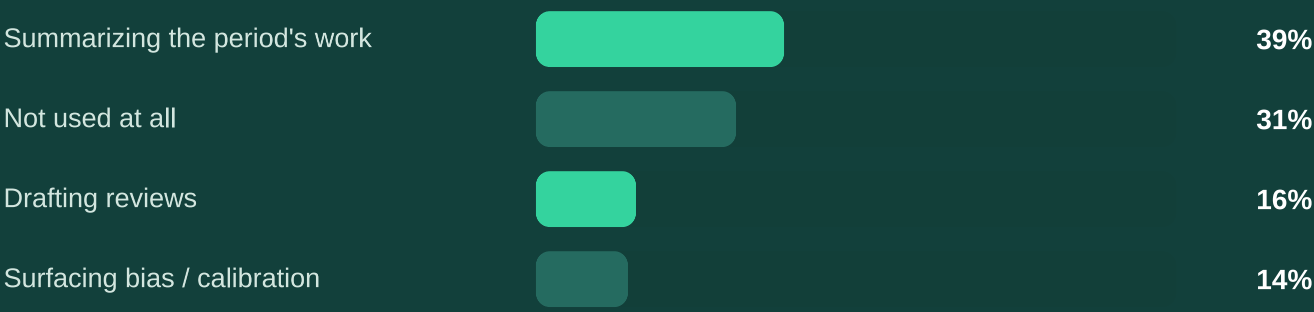
**54%**

Already use AI in the review process

Summarizing a period it never saw

AI has arrived in the performance process — most often to summarize the period's work or draft the review. But AI can only summarize what it can see, and in most organizations what it can see is thin.

## Where AI is used today




> *The most common use of AI — summarizing the period — is the one most exposed to the confidence gap. Among leaders who use AI this way, only 10% run performance on a goal tool and 47% don't capture progress continuously at all. Nearly half are asking AI to summarize a period with no real record behind it — so it does what a manager does from memory: reconstruct, smooth, and invent the specifics.*

Leaders name data privacy as their top AI concern (37%). Far fewer — just 11% — name the more fundamental one: that AI has no goal record to draw from and invents specifics to fill the gap. Privacy is a policy problem you can govern. A missing record is a foundation problem no model compensates for.

SECTION 6

# What It Runs On

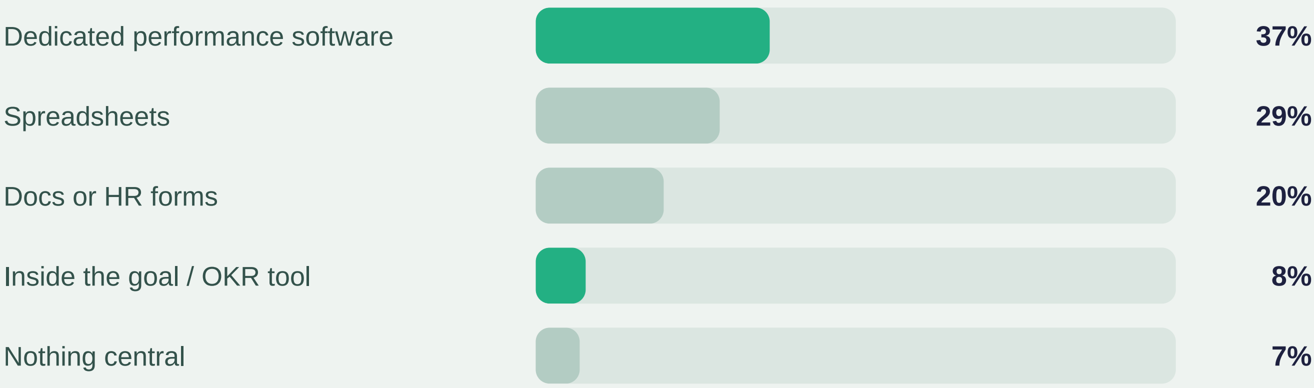
**8%**

Run performance inside their goal tool

The infrastructure behind the gap

Every gap in this report traces to one structural fact: for 92% of organizations, performance runs on infrastructure with no live connection to the goals. The record and the rating live apart because the tools that hold them are apart.

## Where performance actually runs




> *A spreadsheet doesn't remember what happened in month one. A review form doesn't track a goal between cycles. The tools most organizations use are built to capture the rating — not to hold the record that justifies it. That division is the confidence gap in physical form.*

Cadence widens it further: 49% review twice a year or less often — the interval over which early-cycle work reliably disappears from memory.

PROFILE

# What Best Looks Like

A data-informed profile of organizations where the rating and the record are the same thing — the conditions that close the gap between confidence in the system and evidence behind the score.

| | |
|---|---|
| **Recorded** | Goal progress is captured continuously in a live system, not reconstructed before a review. The full period is visible, not just the last few weeks. |
| **Defensible** | Every rating traces to specific goal outcomes. A challenged score resolves by pulling the record — not relitigating memory. |
| **Unbiased by design** | Recency and proximity recede because the score rests on a time-stamped record. Early work and remote work are visible whether or not they're top of mind. |
| **Connected** | The goal cycle and the review cycle are one process on one calendar. At review time the goals are in the room, not fetched from another system. |
| **AI-ready** | AI summarizes and flags from a real, connected goal record — surfacing what happened rather than inventing what it can't see. |
| **Integrated** | Performance runs where the goals live. The record that justifies a rating and the rating itself sit in the same place. |

The distance between this profile and where most organizations sit isn't a capability gap. It's a systems gap — and systems can be fixed.

SELF-ASSESSMENT

# Where Does Your Organization Stand?

Rate yourself honestly across six dimensions. 1 = early stage, 5 = high performance.

| DIMENSION | WHAT YOU'RE SCORING | YOUR RATING |
|---|---|---|
| **Recorded** | From goals reconstructed from memory (1) to a live full-period record (5) | 1 2 3 4 5 |
| **Defensible** | From ratings no one can substantiate (1) to every score traceable to goal evidence (5) | 1 2 3 4 5 |
| **Unbiased** | From recency & proximity driving scores (1) to a time-stamped record driving them (5) | 1 2 3 4 5 |
| **Connected** | From goals & reviews on separate calendars (1) to one integrated cycle (5) | 1 2 3 4 5 |
| **AI-ready** | From AI inventing specifics (1) to AI summarizing a real goal record (5) | 1 2 3 4 5 |
| **Integrated** | From performance on spreadsheets & forms (1) to performance where the goals live (5) | 1 2 3 4 5 |

Total /30

**25–30 High Performance** — the rating and the record are the same thing.
**16–24 Developing** — clear leverage points in your lowest scores.
**Below 16 At Risk** — confidence is running ahead of evidence.

ACTION PLAN

# Four Moves, In Order of Impact

Data-backed. Prioritized. Starting next quarter.

## 1

HIGHEST IMPACT · THE RECORD

### Make the goal record continuous, not reconstructed.

Half of reviews are written from a partial memory of the period. Capture goal progress as it happens, so the full period — not just the last few weeks — is on the record when a rating gets written. You can't defend a score you had to reconstruct.

## 2

QUICK WIN · DEFENSIBILITY

### Require goal evidence for every rating.

Only 27% say nearly all managers could defend a rating. Set the standard that every score points to specific goal outcomes — consistent, not punitive — so a challenge resolves by reading the record, not relitigating memory.

## 3

STRUCTURAL · CONNECTION

### Put the goals and the review on one calendar.

61% run the two cycles separately, so the goals aren't in the room when the rating is written. Integrate them: one cycle, one place, goals visible at review time. This is what makes moves 1 and 2 hold without constant effort.

## 4

INFRASTRUCTURE · TOOLING

### Run performance where the goals live.

92% run performance on infrastructure disconnected from the goals — which is why the record and the rating drift apart. Software that holds both makes recording, defending and connecting structural, not dependent on someone updating a file.

CONCLUSION

# Closing the Confidence Gap

This report set out to answer one question: if ratings are trusted, why can so few be proven? Confidence in the system runs well ahead of the evidence behind any individual score — and the two drift apart for a structural reason, not a human one.

The gap shows up the same way four times. Ratings are trusted but undefendable — only 27% could substantiate every score. They're shaped by memory as much as record — 45% admit something other than goal delivery moves the number. The goals and the review run on separate calendars — 61% keep them apart. And AI is entering faster than the record it needs — 54% use it, 8% give it something real to read.

**66%**

Trust ratings are accurate

**27%**

Could defend every one

**92%**

Run performance apart from goals

None of the fixes are exotic. Record the full period, make every rating traceable, put the goals and the review on one calendar, and run performance where the goals live. These are operating disciplines, not features — and the organizations that run them don't have more confident ratings. They have ratings they can prove.

METHODOLOGY

# About This Research

An independent survey of how performance ratings actually get made.

The Confidence Gap: 2026 Benchmark Report is based on responses from 230 People and HR leaders — those responsible for or directly involved in the performance review process — at technology companies of 50 to 200 employees. All confirmed active involvement before completing the survey. Questions were framed to make ordinary shortcomings feel normal to admit, reducing the social-desirability bias that suppresses honest answers from process owners.

| FIELD | DETAIL |
|---|---|
| Total respondents | **230** |
| Respondent role | Heads of People, HRBPs, People Ops leaders |
| Company size | 50–200 employees |
| Industry | Technology sector |
| Survey period | 2026 |

## A note on the data

All findings are self-reported and reflect respondents' own assessment of their performance and goal processes. Results describe association, not causation. No control group was used, and outcomes vary by organization size, implementation quality, and leadership commitment. No OKRs Tool customers were included in the sample. Because respondents own the processes they're describing, the central finding — the gap between stated confidence and stated defensibility — is reported by the same person on both sides, which makes it resistant to outside-observer bias. We've been deliberately transparent about these limits because an honest number is more valuable than an impressive one.

ABOUT

# OKRs Tool

*Goal-setting built for growing companies that want performance ratings they can actually prove.*

OKRs Tool is a goal-setting platform built for growing companies — without the bloat of enterprise software or the chaos of spreadsheets. Every feature exists to make the goal record the same thing as the performance record, so a rating can always be traced to what actually happened.

- **Real-Time Dashboards**
  Every objective, key result and owner in one place, updated live — the full-period record a fair rating needs.
- **Continuous Check-Ins**
  Lightweight weekly updates capture progress as it happens, so early-cycle work stays on the record.
- **Goals in the Review Room**
  Goal history and how it tracked sit right where the rating gets made — one calendar, nothing to fetch.
- **AI Goal Assistant**
  Summaries and flags drawn from a real, connected goal record — surfacing what happened, not inventing it.

**Build performance ratings you can prove.**

**Start Free at OKRsTool.com →**