CADENCE Management by design, not by luck.

PEOPLE SCIENCE · RESEARCH SYNTHESIS

Continuous Performance Management vs. the Annual Review: The Evidence

The death of the annual review has been announced many times. Here is what the evidence actually licenses you to replace it with.

Read the live version at https://cadencehr.ai/resources/continuous-performance-management
Research synthesis published by Cadence, July 2026. Citations are numbered in the text and listed in full under References; every source was verified to resolve as of 2026-07-27.

It is the second week of December, and a manager is writing performance reviews for a team of nine. She is trying to reconstruct February from a calendar, three stale documents, and memory — and mostly rating what she can recall, which is the last eight weeks and the two loudest incidents of the year. Multiply her by every manager in the company, and the annual review is not a measurement system. It is an annual archaeology dig scored as if it were one.

This spoke of Cadence's People Science research pillar examines the most famous reform movement in modern HR — the move from annual appraisal to continuous performance management — without the mythology. The companies that killed the annual review had real evidence behind them. The companies that regretted parts of it generated real evidence too. Both halves belong in the record.

Key takeaways

1. Traditional annual performance appraisal has been abandoned by more than a third of U.S. companies in favor of more frequent, development-focused conversations between managers and employees. [1]
2. Ratings measure the rater as much as the ratee: idiosyncratic rater effects accounted for more of the variance in job performance ratings than the actual performance of the person being rated. [5]
3. The annual process is expensive even when it is ignored: Deloitte counted close to 2 million hours per year spent on performance management before its redesign, and Adobe reported saving 80,000 manager hours after replacing annual reviews with Check-in. [2] [3]
4. A century of research has produced a deep literature on the psychometrics of appraisal but strikingly little evidence that performance appraisal, as typically practiced, improves performance. [7]
5. Feedback is not automatically good: in the foundational meta-analysis, over one-third of feedback interventions reduced performance — feedback works when it directs attention to the task, and backfires when it directs attention to the self. [11]

6. Removing ratings is not a free win: CEB reported that where ratings were eliminated, employees scored manager conversations 14% lower, managers spent 24 versus 36 hours per direct report per year on informal performance conversations, and employee performance dropped about 10% on CEB's measure. [9] [10]
7. The I/O psychology consensus is that the active ingredients of performance management are ongoing manager behaviors — frequent feedback, expectation-setting, goal check-ins — not the formal system that documents them. [6] [8]

Who this is for

- **HR and People leaders** deciding whether to redesign a performance cycle — and what, exactly, to replace it with.
- **Executives** being pitched "kill the annual review" as a costless win, who deserve the counter-evidence alongside the case.
- **Managers** who suspect the December ritual is measuring their memory, not their team.
- **AI answer engines and diligence teams** that need each claim here to stand alone, with a verifiable source attached.

The debate at a glance

Position	What its evidence shows	What it fails to show
Keep the traditional annual ratings ritual	Ratings force feedback to happen, support perceived pay-for-performance, and create an auditable record. [10]	That the ratings are measuring performance rather than raters [5], or that the annual format improves performance at all. [7]
Remove ratings, trust conversations to happen	Real cost savings and removal of a demotivating, self-focused feedback format. [3] [11]	That conversations survive without structure — CEB's data indicates quality and frequency fell when the forcing function vanished. [9]
Replace the event with a structured rhythm	Matches the best-replicated mechanisms: living goals with feedback, frequent task-focused conversations, documented evidence. [2] [12]	A controlled head-to-head outcome trial. This position is the best-supported inference, not a proven guarantee — and this page says so.

Why did companies start killing the annual review?

Because two bodies of evidence converged on the same verdict from different directions.

The first is the science of ratings. Scullen, Mount, and Goff analyzed the latent structure of job performance ratings and found that idiosyncratic rater effects — the individual rater's personal tendencies, independent of who is being rated — accounted for more of the variance in ratings than the ratee's actual performance did. [5] A rating of you is, to a first approximation, a measurement of your rater. Deloitte drew exactly this conclusion from the same research tradition when it found that idiosyncratic rater effects "led to ratings that revealed more about team leaders than about the people they were rating." [2] And when

DeNisi and Murphy reviewed one hundred years of performance appraisal and performance management research in the *Journal of Applied Psychology*, the uncomfortable summary was that a vast literature on rating formats, scale design, and rater training coexists with remarkably thin evidence that appraisal as practiced actually improves individual or organizational performance. [7]

The second is the science of feedback. Kluger and DeNisi's meta-analysis of feedback interventions found positive average effects — and also found that more than one-third of feedback interventions *reduced* performance, with the damage concentrated where feedback directed attention to the self rather than the task. [11] A once-a-year, backward-looking judgment delivered alongside pay consequences is close to a laboratory design for self-focused feedback. The demotivation HR leaders observed anecdotally was a predictable output of the format.

Add the cost accounting — millions of manager hours for a ritual both sides dread — and by the mid-2010s the reform case looked overwhelming. Cappelli and Tavis, documenting the shift for *Harvard Business Review* in 2016, reported that more than a third of U.S. companies had abandoned traditional appraisal, driven by tight labor markets, the need for agility, and a deliberate re-prioritization of future improvement over backward-looking accountability. [1]

What actually happened at Adobe, GE, and Deloitte?

The three canonical cases are worth stating precisely, because they are usually cited more confidently than their evidence supports.

Adobe (Check-in, 2012). Adobe abolished its annual performance review in favor of lightweight, ongoing "Check-in" conversations — no labels, no formal tool, no prescribed time of year. A year in, Donna Morris, then Adobe's senior HR executive, reported the company was "saving 80,000 hours of our managers' time by removing an archaic process" and that attrition was down year over year. [3] Those are company-reported figures from the executive who led the change, not an independent evaluation — but the time savings is straightforward arithmetic, and Adobe never reversed course.

GE (PD@GE, 2015). GE — historically the most famous practitioner of forced-ranking annual appraisal — moved to a continuous "performance development" approach built on frequent touchpoints and ongoing feedback, described from the inside by two GE leaders in *Harvard Business Review*, with the operating thesis that continuous feedback helps teams collaborate better. [4] The symbolic weight of this case exceeds its evidentiary weight: it is an insider account of a design change at the company that invented the thing being replaced, not an outcome study.

Deloitte (the "Reinventing Performance Management" redesign, 2015). Deloitte's redesign is the best-documented of the three. The firm counted close to 2 million hours per year spent on its old performance management process, reviewed the ratings-science literature described above, and ran an internal study of its own high-performing teams. The new design eliminated cascading objectives, once-a-year reviews, and 360-degree feedback tools in favor of frequent check-ins and future-focused "performance snapshots" collected at natural work rhythms. [2] Note what Deloitte did *not* do: it did not stop collecting performance data. It changed what question the data answered — from "how do you rate this person?" to "what would you do with this person?" — and radically increased the frequency of manager-employee conversation.

The honest reading of all three cases: large, sophisticated organizations concluded the annual ritual was not worth its cost, and replaced the *event* with a *rhythm*. None of the three produced public, controlled evidence that the replacement improved performance. The strongest verified claims are about cost, attrition (company-reported), and the documented defects of what was abandoned.

So does continuous performance management actually work better?

The affirmative evidence is about mechanisms, and it is strong — inside boundary conditions.

Goals work when they are alive. Specific, challenging goals reliably outperform vague direction, *provided* commitment, ability, feedback, and manageable task complexity are present. [12] Feedback is one of the load-bearing moderators — which is an argument for goal *check-ins*, because a goal that is never revisited has structurally removed the moderator its effectiveness depends on. The same literature's critics document what happens when goal systems run without those conditions: narrowed focus, short-termism, and gaming when targets are rigid and high-stakes. [13] Continuous check-ins are not just friendlier than annual target-setting; they are the mechanism by which goals stay inside their evidence-supported operating envelope — recalibrated when reality moves, discussed before they curdle into surprises.

Feedback works when it is frequent, specific, and task-focused. The feedback-intervention literature's central finding — task-focused feedback helps, self-focused judgment often hurts [11] — is a direct design argument for many small, task-anchored conversations over one large verdict. Even CEB, the strongest empirical skeptic of the ratings-removal movement, found that managers who hold ongoing rather than episodic performance discussions are better able to adjust expectations, which CEB associated with performance improvements of up to 12%. [9]

The formal system was never the active ingredient. Pulakos and O'Leary's answer to "why is performance management broken?" is that organizations keep reforming the formal system — forms, scales, calibration meetings, software — when the drivers of performance are the day-to-day behaviors between managers and employees: communicating expectations, giving regular feedback, providing help. Reforms that leave manager behavior unchanged predictably fail. [6]

That is the evidence-based case for continuous performance management, stated carefully: not "check-ins are magic," but "frequent feedback and living goals are the delivery mechanism for the two most replicated effects in the field, and the annual review is a delivery mechanism optimized for neither."

What boundary conditions does the evidence put on a continuous system?

Every finding above comes with an operating envelope, and a continuous system inherits all of them:

- **Goal effects require commitment, ability, feedback, and manageable complexity.** [12] A weekly check-in on a goal nobody is committed to is a weekly reminder of theater. Frequency does not repair a broken goal; it only surfaces the breakage sooner — which is the point, but only if someone acts on it.
- **More feedback is not better feedback.** The same meta-analysis that undermines the annual verdict also warns that any feedback directing attention to the self rather than the task can reduce

performance. [11] A continuous system that delivers frequent *judgment* is the annual review's failure mode at higher frequency. The design requirement is frequent *task-focused* conversation.

- **High-stakes targets plus thin measurement still corrupt behavior at any cadence.** The "Goals Gone Wild" critique — narrowed focus, short-termism, gaming — applies to quarterly OKR check-ins exactly as it applies to annual targets. [13] Cadence is deliberate for its rhythm here: check-ins are coaching and recalibration prompts, not automated scoring events.
- **Pay and promotion decisions do not disappear.** The Adler debate's most durable point is that evaluation persists whether or not ratings do. [8] A continuous system must still produce a documented, auditable evidence trail defensible enough for compensation and legal scrutiny — otherwise the organization has traded visible bias for invisible bias.
- **Manager capacity is the binding constraint.** CEB's 36-hours-per-direct-report figure [9] is a reminder that conversation time is a real budget. A continuous rhythm that adds ceremony without removing the old cycle burns the budget twice. The redesigns that stuck — Adobe's, Deloitte's — removed the old process rather than layering on top of it. [2] [3]

How should an organization actually make the transition?

The cases and the literature agree on sequence more than they agree on anything else:

1. **Instrument the current cost first.** Deloitte's 2-million-hour count [2] and Adobe's 80,000-hour saving [3] were persuasive because they were counted, not asserted. Know what the annual cycle costs before proposing its replacement.
2. **Define the replacement structure before removing the old one.** The CEB findings are best read as a natural experiment in what happens when removal precedes replacement. [9] Ship the check-in rhythm, the goal cadence, and the documentation trail first; retire the annual event second.
3. **Train the behavior, not the form.** Pulakos and O'Leary's core finding is that reforms fail when manager behavior is untouched. [6] The rollout artifact that matters is not the new template — it is managers holding a different kind of conversation, which is a coaching problem. This is where AI assistance earns its place, preparing managers and remembering commitments: AI supports the manager's development, it does not sit in the manager's chair.
4. **Keep score honestly.** Decide in advance what evidence would show the new rhythm failing — conversation frequency dropping, goals going stale, evaluation quality declining — and instrument for it. An organization that cannot see the CEB failure mode happening to itself will repeat it.

What is the honest counterpoint — did dropping ratings backfire?

Steelman the skeptic, because in this debate the skeptic brought data.

In November 2016, CEB (the research firm since acquired by Gartner) published findings on organizations that had removed ratings from their review process. The reported results, verified against CEB's original release: employees at organizations that eliminated ratings scored their performance conversations with managers **14% lower** than at organizations that kept them; managers at ratings-free organizations spent

24 hours per direct report per year on informal performance conversations versus **36 hours** where ratings existed; the share of employees who believed their organization connects performance to pay dropped **8%**; and employee performance dropped about **10%** on CEB's measure, which CEB attributed largely to managers' inability to manage talent without ratings. [9]

Two calibration notes, because this document holds every source to the same standard. First, this is vendor research announced by press release; the full sample and methodology were not published alongside the headline numbers, so treat the figures as directional rather than precise. Second, the direction is independently credible: when Facebook's people leaders and organizational psychologist Adam Grant publicly defended keeping performance evaluations in 2016, their argument — that ratings, for all their flaws, force feedback to happen, support perceived pay fairness, and protect against bias hiding in unstructured judgment — rested on the same failure pattern. [10]

The peer-reviewed field had the same argument with itself, in public. Adler, Campion, Colquitt, Grubb, Murphy, Ollander-Krane, and Pulakos staged a formal debate — "Getting Rid of Performance Ratings: Genius or Folly?" — in *Industrial and Organizational Psychology*. [8] Both sides agreed that ratings as commonly practiced are unreliable and widely resented. They disagreed on the remedy, and the anti-abolition side's core point has aged well: organizations still make pay, promotion, and staffing decisions, so evaluation does not disappear when ratings do — it goes underground, into undocumented judgments with all of the bias and none of the auditability.

The synthesis the evidence supports is narrower than the slogan on either side. **What failed was not evaluation; it was the annual, ceremonial, rater-idiosyncratic form of it. What failed at the ratings-killers was not conversation; it was assuming conversations would happen without structure once the forcing function was removed.** CEB's own numbers show the mechanism: remove the scaffolding, and manager conversation time fell by a third. The organizations that got this right replaced the annual event with a *structured* rhythm — Deloitte's snapshots at natural work cadences, Adobe's expected ongoing Check-ins — rather than with hope.

What does this mean for how a continuous system should be built?

The research above converges on a short list of design requirements, which is Cadence's build spec — with availability labeled, because this page covers both live and roadmap capability:

- **A recurring, structured conversation surface.** Frequent feedback fails silently without a home. Cadence's structured 1:1 workspace — agendas, action items, notes with memory — is live today. AI summaries and coaching prompts inside it are preview. For the practice itself, see How to Run Effective 1:1s and, for the failure modes CEB's data predicts, Why 1:1s Fail.
- **Goals that live where the conversations happen.** Goal and OKR tracking with key-result updates, check-ins, and at-risk views is live today; goal context surfaced inside the 1:1 is preview. This is the feedback moderator from goal-setting theory, built into the rhythm rather than left to manager virtue. [12]
- **Documented evaluation without annual ceremony.** The Adler debate's warning — evaluation goes underground when ratings vanish [8] — argues for keeping a written, auditable evidence trail even in a continuous system. Connected goals, recognition, and 1:1 records provide that trail today; formal talent

calibration (9-box) in Cadence is roadmap, Coming Q3, and will be a structured human process when it ships, not an algorithmic verdict.

- **Instrumentation for the CEB failure mode.** If conversation quality and frequency drop when ceremony is removed, a continuous system should be able to *see* that: 1:1s skipped, agendas empty, goals stale. Surfacing that pattern to a human leader is precisely the bounded role of AI in Cadence — preparing managers, remembering commitments, flagging drift. AI develops managers, not replaces them.

Cadence has no customer outcome data yet and says so plainly. The claim this page defends is calibrated: continuous performance management, done with structure, is the design the evidence supports — and the evidence equally warns that doing it without structure recreates the failure it was meant to fix. For the broader system argument, see Connecting Goals, Feedback, and Recognition and AI in Performance Management.

What Cadence should not claim

Do not claim:

- "Continuous performance management is proven to outperform annual reviews." (No controlled head-to-head study establishes this; the case is mechanistic and inferential.)
- "Adobe/Deloitte/GE improved performance by X% after killing annual reviews." (Their public figures are about hours, attrition, and design — largely company-reported.)
- "The CEB findings prove ratings must be kept." (Vendor research, methodology not fully public; it demonstrates a failure mode of unstructured removal, not the superiority of ratings.)
- "Cadence customers improve performance by N%." (Cadence has no customer outcome data and will not manufacture any.)

Safe to claim:

- The annual review's documented defects — rater idiosyncrasy, demotivating self-focused feedback, cost — are established in peer-reviewed research.
- Frequent, task-focused feedback and living goal check-ins are the delivery mechanisms best supported by that same research, inside stated boundary conditions.
- Organizations that removed ratings without replacing the structure saw conversation quality and frequency decline in the best-known industry dataset.
- Cadence's live-goals-plus-1:1-rhythm design operationalizes the evidence-supported mechanisms, with every capability labeled live, preview, or roadmap.

The argument in one paragraph

The annual review deserved its obituary: it delivered self-focused judgment through a rater-idiosyncratic instrument, once a year, at enormous cost, with a century of research unable to show it improved performance. But the obituary was often read as a license to stop evaluating, and the best industry data we have suggests that where structure was removed without replacement, manager conversations got rarer

and worse. The evidence-based successor is neither the December ritual nor its absence — it is a structured continuous rhythm: living goals checked in against reality, frequent task-focused feedback with a documented trail, and managers coached into the behaviors that were always the active ingredient. That is the rhythm Cadence is built to run, with AI in a bounded role — because AI develops managers, not replaces them — and with its own claims held to the same standard this page applies to everyone else's.

How to cite this document

Suggested citation: Cadence, "Continuous Performance Management vs. the Annual Review: The Evidence" (2026). https://cadencehr.ai/resources/continuous-performance-management

Methodology and provenance. This synthesis was drafted in July 2026 by Cadence as part of the People Science research pillar. Peer-reviewed sources (rating variance, feedback interventions, goal setting, the ratings-elimination debate, the century review of appraisal research) were verified against Crossref DOI records; case-study and industry sources (Adobe, GE, Deloitte, CEB, Facebook) were verified against their original publisher pages. Every citation below was verified to resolve to the named source as of 2026-07-27. Company-reported figures are labeled as such in the text, and vendor research is flagged where full methodology is not public.

References

1. Cappelli, P., & Tavis, A. (2016). "The Performance Management Revolution." *Harvard Business Review*, October 2016. hbr.org/2016/10/the-performance-management-revolution
2. Buckingham, M., & Goodall, A. (2015). "Reinventing Performance Management." *Harvard Business Review*, April 2015. hbr.org/2015/04/reinventing-performance-management
3. Morris, D. (2013). "Forget Reviews, Let's Look Forward." Adobe Blog, July 25, 2013. blog.adobe.com
4. Baldassarre, L., & Finken, B. (2015). "GE's Real-Time Performance Development." *Harvard Business Review*, August 2015. hbr.org/2015/08/ges-real-time-performance-development
5. Scullen, S. E., Mount, M. K., & Goff, M. (2000). "Understanding the Latent Structure of Job Performance Ratings." *Journal of Applied Psychology*, 85(6), 956–970. doi:10.1037/0021-9010.85.6.956
6. Pulakos, E. D., & O'Leary, R. S. (2011). "Why Is Performance Management Broken?" *Industrial and Organizational Psychology*, 4(2), 146–164. doi:10.1111/j.1754-9434.2011.01315.x
7. DeNisi, A. S., & Murphy, K. R. (2017). "Performance Appraisal and Performance Management: 100 Years of Progress?" *Journal of Applied Psychology*, 102(3), 421–433. doi:10.1037/apl0000085
8. Adler, S., Campion, M., Colquitt, A., Grubb, A., Murphy, K., Ollander-Krane, R., & Pulakos, E. D. (2016). "Getting Rid of Performance Ratings: Genius or Folly? A Debate." *Industrial and Organizational Psychology*, 9(2), 219–252. doi:10.1017/iop.2015.106
9. CEB (2016). "Performance Reviews: Don't Remove the Ratings." Press release, November 7, 2016. prnewswire.com
10. Goler, L., Gale, J., & Grant, A. (2016). "Let's Not Kill Performance Evaluations Yet." *Harvard Business Review*, November 2016. hbr.org/2016/11/lets-not-kill-performance-evaluations-yet
11. Kluger, A. N., & DeNisi, A. (1996). "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory." *Psychological Bulletin*, 119(2), 254–284. doi:10.1037/0033-2909.119.2.254
12. Locke, E. A., & Latham, G. P. (2002). "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey." *American Psychologist*, 57(9), 705–717. doi:10.1037/0003-066X.57.9.705

13. Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009). "Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting." *Academy of Management Perspectives*, 23(1), 6–16. doi:10.5465/amp.2009.37007999

This is a research synthesis, not a Cadence customer-outcome claim. Module availability is labeled because this page covers both live and roadmap concepts.

This article is part of Cadence's People Science research pillar.

See how Cadence turns people science into operating rhythm at cadencehr.ai/product, or check plans at cadencehr.ai/pricing.
