



Lisbon School
of Economics
& Management
Universidade de Lisboa

MESTRADO
GESTÃO DE RECURSOS HUMANOS

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

**ERROS RECORRENTES E RECOMENDAÇÕES NA AVALIAÇÃO DE
DESEMPENHO: UMA REVISÃO SISTEMÁTICA DA LITERATURA
(2010-2025)**

MICKÉYAS PEREIRA DE PAULA LEITE

JULHO – 2026

MESTRADO
GESTÃO DE RECURSOS HUMANOS

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

**ERROS RECORRENTES E RECOMENDAÇÕES NA AVALIAÇÃO DE
DESEMPENHO: UMA REVISÃO SISTEMÁTICA DA LITERATURA
(2010-2025)**

MICKÉYAS PEREIRA DE PAULA LEITE

ORIENTAÇÃO:

PROFESSOR DOUTOR PAULO LOPES HENRIQUES

JULHO – 2026

AGRADECIMENTOS

A conclusão deste Trabalho Final de Mestrado representa muito mais do que o cumprimento de uma exigência acadêmica. Representa o fim de uma longa caminhada, rigorosa e, por várias vezes, silenciosamente difícil. Houve momentos de cansaço, incerteza, distância, renúncia e solidão. Houve dias em que continuar parecia mais uma obrigação do que uma escolha. Ainda assim, cheguei até aqui porque, ao longo desse caminho, encontrei alicerces que me sustentaram quando a vontade, por si só, talvez já não fosse o bastante.

Agradeço, em primeiro lugar, a Deus, por ser a minha direção, por me guardar nos momentos de fragilidade e por me permitir chegar até aqui. Em muitas ocasiões, quando as respostas não eram tão claras e o caminho parecia muito pesado, foi na fé que encontrei paz para continuar.

Ao Professor Doutor Paulo Lopes Henriques, o meu mais sincero agradecimento pela orientação, pela disponibilidade e pela forma criteriosa com que acompanhou este trabalho. A sua capacidade intelectual, a clareza das suas observações e suas rigorosas intervenções foram fundamentais para que este estudo chegasse à melhor forma. O professor Paulo indicou caminhos com precisão, exigência e firmeza, fazendo com que este trabalho fosse concluído com maior consistência acadêmica.

Ao Corpo de Bombeiros Militar do Distrito Federal, instituição pública militar brasileira à qual pertenço e que atua na proteção da vida, do património e da sociedade no Distrito Federal, agradeço pela oportunidade de viver esta experiência de formação e aprendizagem. Levo comigo a responsabilidade de retribuir à instituição, com dedicação, serviço e compromisso, parte daquilo que esta etapa me permitiu construir.

Às minhas filhas, Bianca e Júlia, agradeço de forma muito especial. Elas foram a razão mais profunda para continuar, mesmo nos dias em que a saudade de brincar, ou somente estar com elas, pesava demais. Entre a tristeza pela distância e pela falta que tantas vezes senti, havia sempre sorrisos pela simples existência delas. Foram saudade e força, fragilidade e coragem, ausência e presença ao mesmo tempo. A elas dedico parte essencial desta conquista, porque, mesmo quando eu estava longe, eram elas que me lembravam por que valia a pena seguir.

Ao meu papai, Paulo, agradeço por sempre me incentivar a estudar e por acreditar no meu potencial. O seu apoio e a sua confiança foram, ao longo da minha vida, uma lembrança constante de que o conhecimento abre caminhos e transforma destinos.

À minha mamãe, Ronilda, agradeço de forma muito especial. Mais do que mãe, foi amiga, presença constante, abrigo e apoio irrestrito. O seu amor incondicional atravessou o oceano, acolheu cansaços e sustentou-me em muitos momentos em que eu não consegui demonstrar o peso que carregava. Sem a sua presença, mesmo quando fisicamente distante, esta caminhada teria sido muito mais difícil.

Ao meu irmão, Japhet, agradeço pelas palavras constantes de apoio, pela sensatez e pela tranquilidade que tantas vezes me ajudaram a reorganizar pensamentos e a recuperar a direção. Agradeço também pelo exemplo acadêmico e profissional que ele sempre foi para mim, e pela forma como sempre me transmitiu segurança, maturidade e confiança.

À minha cunhada, Suellen Steffany, e aos meus sobrinhos, Davi e Lucas, Asaph e Isaac, agradeço pelo carinho, pela presença familiar e por fazerem parte dessa rede de afeto que tornou a caminhada menos pesada. Em períodos de exigência e afastamento, a família foi também um lembrete de pertencimento e de cuidado para além das obrigações.

À minha tia Virgínia Palazzo, agradeço por todo o apoio, pelas orações, pelo amor e pelas palavras de incentivo. A sua fé e o seu cuidado estiveram comigo nos momentos em que precisei renovar as forças.

Por fim, agradeço a todos aqueles que, de alguma forma, fizeram parte deste percurso. Este trabalho leva o meu nome, mas não foi construído apenas por mim. Ele carrega também a força, a fé, o amor e a confiança de pessoas e instituições que me sustentaram ao longo do caminho. Sem esses alicerces, muito provavelmente, esta etapa não teria sido concluída.

RESUMO

A avaliação de desempenho é fundamental na gestão de recursos humanos, porém ainda apresenta muitos erros, preconceitos e falhas estruturais que comprometem a justiça, a precisão e a utilidade dos julgamentos de avaliação. Este estudo analisa os desafios mais frequentes enfrentados pelos sistemas de avaliação de desempenho, bem como as soluções propostas pela ciência para superá-los. Para isso, foi realizada uma revisão sistemática da literatura, seguindo as diretrizes do PRISMA 2020 abrangendo estudos publicados entre 2010 e 2025 das bases EBSCO, *Scopus* e *Web of Science*. Seguindo os critérios de elegibilidade, o corpus final foi constituído por 37 artigos, sendo analisados com o uso de uma matriz de extração e síntese temática. Os resultados sugerem que as questões relacionadas à avaliação de desempenho vão além do julgamento individual do avaliador, mas também decorrem da interação entre subjetividade, falhas de desenho e implementação, critérios pouco claros, feedback inadequado, documentação insuficiente, falta de calibração e fatores culturais, relacionais e institucionais. Os vieses clássicos, como o efeito *halo*, o efeito *horn*, a leniência, a severidade, a recência e a tendência central, continuam relevantes, entretanto, tendem a se intensificar em sistemas pouco estruturados e com práticas de avaliação inconsistentes. É importante notar que as recomendações mais recorrentes indicam a necessidade de esclarecimento de critérios, esclarecimento de metas observáveis, treinamento e calibração de avaliadores, feedback contínuo, participação dos avaliados, transparência, registo de evidências e alinhamento estratégico do processo de avaliação com as decisões de gestão de recursos humanos. Observa-se que limitar erros e preconceitos requer uma abordagem a nível de sistema para ser capaz de articular pessoas, processos, instrumentos, cultura organizacional e propósito institucional, tornando a avaliação de desempenho um processo mais justo, rastreável e orientado para o desenvolvimento.

Palavras-chave: avaliação de desempenho; erros de avaliação; vieses cognitivos; revisão sistemática da literatura; gestão de recursos humanos.

ABSTRACT

Performance evaluation is fundamental to human resource management, yet it still contains many errors, biases, and structural flaws that compromise the fairness, accuracy, and usefulness of evaluation judgments. This study analyzes the most common challenges faced by performance evaluation systems, as well as the solutions proposed by the scientific community to overcome them. To this end, a systematic literature review was conducted following the PRISMA 2020 guidelines, covering studies published between 2010 and 2025 from the EBSCO, Scopus, and Web of Science databases. Based on the eligibility criteria, the final corpus consisted of 37 articles, which were analyzed using a data extraction and thematic synthesis matrix. The results suggest that issues related to performance evaluation are not limited solely to the evaluator's individual judgment but arise from the interaction between subjectivity, design and implementation flaws, unclear criteria, inadequate feedback, insufficient documentation, lack of calibration, and cultural, relational, and institutional factors. Classic biases, such as the halo effect, the horn effect, leniency, severity, recency, and central tendency, remain relevant; however, they tend to intensify in unstructured systems and inconsistent evaluation practices. It is important to note that the most common recommendations point to the need for clear criteria, observable goals, training and calibration of evaluators, continuous feedback, employee participation, transparency, documentation of evidence, and strategic alignment of the evaluation process with human resources management decisions. It is observed that limiting errors and biases requires a system-level approach to effectively integrate people, processes, tools, organizational culture, and institutional purpose, making performance evaluation a fairer, more traceable, and development-oriented process.

Keywords: performance evaluation; evaluation errors; cognitive biases; systematic literature review; human resource management.

ÍNDICE

1.	INTRODUÇÃO	11
2.	REVISÃO DA LITERATURA	14
2.1.	Avaliação de desempenho: conceitos fundamentais	14
2.2.	Evolução dos sistemas de avaliação e mudanças no trabalho	16
2.3.	Modelos, métodos e critérios de avaliação do desempenho	17
2.4.	Erros e vieses na avaliação de desempenho	18
2.4.1.	<i>Distorções clássicas na avaliação de avaliação</i>	19
2.4.2.	<i>Vieses cognitivos e influências contextuais</i>	20
2.5.	Consequências dos erros na avaliação de desempenho	21
2.6.	Inconsistências e limitações teóricas identificadas	22
3.	METODOLOGIA	22
3.1.	<i>Estratégia de pesquisa</i>	24
3.2.	<i>Justificação da baliza temporal</i>	24
3.3.	<i>Justificação dos idiomas incluídos</i>	25
3.4.	<i>Resultados da pesquisa nas bases de dados</i>	26
3.5.	<i>Crerios de elegibilidade</i>	26
3.6.	<i>Processo de seleção dos estudos</i>	27
3.7.	<i>Fluxograma PRISMA</i>	28
3.8.	<i>Extração e organização dos dados</i>	29
3.9.	<i>Estratégia de análise</i>	30
4.	ANÁLISE DOS RESULTADOS	31
4.1.	<i>Caracterização geral dos estudos incluídos</i>	32
4.2.	<i>Problemas, erros e vieses mais recorrentes</i>	35
4.3.	<i>Articulação entre problemas e mecanismos de mitigação</i>	39
4.4.	<i>Recomendações recorrentes e níveis de análise</i>	43
4.5.	<i>Síntese analítica dos resultados</i>	46
5.	CONCLUSÕES, CONTRIBUTOS, LIMITAÇÕES E INVESTIGAÇÃO FUTURA	47
5.1.	<i>Conclusões principais</i>	48

5.2.	<i>Contributos teóricos e implicações práticas</i>	51
5.3.	<i>Limitações</i>	52
5.4.	<i>Investigação futura</i>	53
5.5.	<i>Considerações finais</i>	54
REFERÊNCIAS BIBLIOGRÁFICAS		55

ÍNDICE DE TABELAS

TABELA I OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL DA ONU RELACIONADOS COM PRÁTICAS DE AVALIAÇÃO DE DESEMPENHO	12
TABELA II ERROS E VIESES CLÁSSICOS NA AVALIAÇÃO DE DESEMPENHO.....	19
TABELA III RESULTADOS DA PESQUISA NAS BASES DE DADOS	26
TABELA IV NATUREZA DOS ESTUDOS INCLUÍDOS.....	33
TABELA V CATEGORIAS ANALÍTICAS MAIS RECORRENTES ASSOCIADAS À AVALIAÇÃO DE DESEMPENHO.....	35
TABELA VI QUADRO ANALÍTICO ENTRE PROBLEMAS RECORRENTES E MECANISMOS DE MITIGAÇÃO	40
TABELA VII SÍNTESE DOS NÍVEIS DE ANÁLISE DAS RECOMENDAÇÕES.....	44

ÍNDICE DE FIGURAS

FIGURA 1 - Fluxograma PRISMA para a Apresentação do Processo de Seleção dos Estudos	29
FIGURA 2 - Gráfico - Distribuição temporal dos artigos.....	32
FIGURA 3 - Gráfico - Abordagem Metodológica Predominante	34
FIGURA 4 - Mapa de coocorrência das categorias codificadas	38
FIGURA 5 - Gráfico - Recomendações recorrentes identificadas na matriz de análise ...	43

1. INTRODUÇÃO

A avaliação de desempenho (AD) é uma prática central na gestão de pessoas, utilizada para compreender como o contributo individual se traduz nos resultados organizacionais. Historicamente, a AD foi concebida como um instrumento retrospectivo, que registava acontecimentos e comportamentos passados, funcionando como um “espelho retrovisor” que documenta trajetórias, mas não define rumos futuros. Contudo, diante das transformações recentes no trabalho, tais como a intensificação da digitalização, a expansão do trabalho híbrido e a crescente complexidade das interações profissionais, essa visão estritamente retrospectiva tornou-se insuficiente.

Em contextos marcados pela velocidade da informação e pela dispersão geográfica das equipas, a AD não pode limitar-se a registar o passado. A lógica da gestão contemporânea exige processos mais dinâmicos, transparentes e orientados ao desenvolvimento contínuo, capazes de apoiar decisões, alinhar expectativas e promover comportamentos que sustentem a missão organizacional.

Processos avaliativos justos e transparentes reforçam a motivação, a produtividade e a confiança nas relações de trabalho. Aguinis (2014, p. 29) destaca que a eficácia dos sistemas avaliativos depende em grande parte das perceções de justiça distributiva, procedimental, interpessoal e informacional, elementos que sustentam a credibilidade do processo. DeNisi e Murphy (2017, p. 426) acrescentam que a avaliação de desempenho suporta um papel central na motivação, no desenvolvimento dos colaboradores e na eficácia organizacional, contribuindo para ambientes de trabalho mais cooperativos e orientados ao desempenho. Estes elementos evidenciam que a AD não é apenas uma ferramenta administrativa, mas um mecanismo que influencia

comportamentos, percepções de justiça e a própria experiência de trabalhar numa organização.

Deste modo, a avaliação de desempenho não se limita a uma ferramenta interna de gestão, pois ultrapassa o plano organizacional e assume também uma dimensão social e ética, na medida em que pode contribuir para organizações mais justas, transparentes e alinhadas com os Objetivos de Desenvolvimento Sustentável. A Agenda 2030 das Nações Unidas estabelece metas relacionadas com o crescimento económico sustentável, a redução das desigualdades e o fortalecimento de instituições eficazes e transparentes (United Nations, 2015). Entre os Objetivos de Desenvolvimento Sustentável (ODS), três estão relacionados diretamente, conforme pode ser visto na Tabela I.

TABELA I
OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL DA ONU RELACIONADOS COM
PRÁTICAS DE AVALIAÇÃO DE DESEMPENHO

ODS	Descrição Resumida	Relação com Avaliação de Desempenho
ODS 8 – Trabalho digno e crescimento económico	Promover crescimento sustentado e emprego digno.	AD justa reforça motivação, produtividade e desenvolvimento do capital humano.
ODS 10 – Redução das desigualdades	Reduzir desigualdades dentro dos países.	AD meritocrática reduz enviesamentos e desigualdades internas.
ODS 16 – Paz, justiça e instituições eficazes	Promover instituições eficazes, transparentes e responsáveis.	AD transparente reforça confiança, responsabilidade e credibilidade institucional.

Nota. Elaboração própria com base na Agenda 2030 para o Desenvolvimento Sustentável (United Nations, 2015).

Apesar desta importância, a literatura revela divergências na definição de conceitos, na compreensão do papel do avaliador e nas práticas recomendadas para garantir processos mais equitativos e eficazes. Estudos como os de Lameque et al. (2023,

pp. 3-4) identificam falta de convergência conceptual, discrepâncias metodológicas e recomendações algumas vezes contraditórias. Além disso, a partir de 2010 houve uma intensificação dos debates sobre justiça organizacional, vieses cognitivos, feedback contínuo e transformações tecnológicas que alteraram significativamente as práticas de avaliação. A literatura recente revela não apenas uma multiplicidade de erros e vieses associados à avaliação de desempenho, mas também um conjunto significativo de propostas de melhoria, que vão desde ajustes técnicos nos instrumentos até mudanças culturais e estratégicas nas organizações.

Neste contexto, e procurando integrar de forma sistemática os contributos dispersos da literatura, a presente investigação procura responder às seguintes questões:

Quais são os erros, vieses e falhas estruturais mais recorrentes identificados nos sistemas de avaliação de desempenho entre 2010 e 2025? Quais conselhos de mitigação são indicados em níveis individual, interpessoal, organizacional e estratégico?

Para operacionalizar esta questão principal, foram definidas as seguintes subquestões de investigação:

1. Quais são os erros e vieses cognitivos mais frequentemente identificados nos sistemas de avaliação de desempenho entre 2010 e 2025?
2. Que falhas processuais, contextuais e estruturais contribuem para a distorção dos julgamentos avaliativos?
3. Que recomendações, práticas ou mecanismos de mitigação são propostos pela literatura científica para reduzir erros e vieses na avaliação de desempenho?
4. Em que níveis de análise (individual, interpessoal, organizacional e estratégico) se concentram as soluções identificadas na literatura?

5. Que contributos, lacunas e tendências emergem da literatura recente relativamente à melhoria dos sistemas de avaliação de desempenho?

Para responder a essas questões, foi adotada uma revisão sistemática da literatura orientada pelo PRISMA 2020, de modo a assegurar rigor, transparência e reprodutibilidade no processo de seleção e análise dos estudos. Assim, a partir da leitura integral dos artigos incluídos, este trabalho procurou clarificar os principais problemas que afetam a avaliação de desempenho e sistematizar as recomendações mais recorrentes da literatura recente, oferecendo uma base conceptual e prática para o desenho de sistemas avaliativos mais justos, precisos e eficazes.

2. REVISÃO DA LITERATURA

2.1. *Avaliação de desempenho: conceitos fundamentais*

A avaliação de desempenho tem raízes mais antigas do que a própria Gestão de Recursos Humanos (GRH). Wilkinson et al. (2017, p. 190) afirmam que práticas avaliativas formais já eram utilizadas muito antes da consolidação da GRH, referindo, entre outros exemplos, o registo sistemático de desempenho na administração chinesa do século III e o uso do *silent monitor* nas fábricas têxteis de Robert Owen no início do século XIX, complementado por anotações anuais nos *books of character*. Estes registos são hoje considerados precursores diretos das práticas modernas de avaliação de desempenho.

Apesar destes antecedentes, a divulgação moderna da avaliação de desempenho ocorre especialmente ao longo do século XX, acompanhando a expansão das práticas formais de gestão. Murphy e Cleveland (1995) descrevem este período como um marco

na institucionalização da AD, momento em que a expressão *performance appraisal* passa a integrar de forma sistemática as rotinas de gestão das organizações norte-americanas.

Já a Gestão de Recursos Humanos, tal como é vista hoje, surge apenas na década de 1980, com a consolidação de modelos que redefiniram o papel estratégico da gestão de pessoas. Embora as práticas de pessoal já existissem, é com os chamados *Matching Models* que a GRH moderna ganha identidade conceptual. O modelo de Michigan, proposto por Fombrun et al. (1984, como citado em Wilkinson et al., 2017, p. 8) enfatiza o alinhamento entre estratégia organizacional e políticas de recursos humanos, enquanto o modelo de Harvard, proposto por Beer et al. (1985, como citado em Wilkinson et al., 2017, p. 8) destaca a importância dos múltiplos *stakeholders* e do desenvolvimento humano como elementos centrais do sistema de gestão. Esta dupla contribuição marca o momento em que a gestão de pessoas deixa de ser predominantemente administrativa e passa a ser vista como um processo estratégico capaz de influenciar o desempenho organizacional.

Neste percurso histórico, a avaliação de desempenho evolui de um instrumento meramente burocrático para um processo que incorpora desenvolvimento contínuo, aprendizagem e alinhamento estratégico. Aguinis (2014, pp. 2-3) aponta que avaliar é um processo orientado para o futuro, promovido por diálogos consistentes sobre expectativas, competências e comportamentos. Armstrong (2015, p. 84) reforça esta perspectiva ao apresentar a gestão de desempenho como um ciclo permanente de planeamento, acompanhamento, feedback e revisão. Segundo Lunenburg (2012, p. 1), o feedback estruturado é fundamental para deixar claro quais os comportamentos esperados e, assim, orientar direções, permitindo corrigir desvios antes que se consolidem em problemas.

Atualmente, existe a ideia de que a avaliação de desempenho é mais do que um registo formal, ou seja, quando estruturada de forma clara e consistente, a AD promove expectativas, favorece a motivação, reforça percepções de justiça e sustenta relações de confiança entre colaboradores e gestores, elementos que influenciam diretamente o empenho e a satisfação no trabalho. Assim, a AD deixa de ser apenas uma técnica de classificação e torna-se papel central na construção de culturas organizacionais baseadas em responsabilidade, aprendizagem e melhoria contínua.

2.2. Evolução dos sistemas de avaliação e mudanças no trabalho

Os sistemas de AD evoluíram significativamente nas últimas décadas, acompanhando transformações sociais, tecnológicas e organizacionais. Modelos tradicionais, centrados em avaliações anuais e classificações predominantemente quantitativas, tornaram-se insuficientes para captar a complexidade do comportamento humano e o ritmo acelerado das mudanças no trabalho.

DeNisi e Murphy (2017, p. 421) mostram que esta evolução vem da necessidade de alinhar avaliação, desenvolvimento e estratégia organizacional, promovendo coerência entre objetivos organizacionais e comportamentos esperados.

Murphy e DeNisi (2023, p. 144) observam que várias organizações têm procurado abandonar os sistemas tradicionais de avaliação de desempenho, influenciadas por tendências recentes e pela percepção de que estes instrumentos apresentam falhas. No entanto, os autores alertam que eliminar as avaliações não resolve os problemas existentes e pode mesmo gerar riscos adicionais, sendo essa decisão comparável a “deitar fora o bebé com a água do banho”. Em vez disso, defendem que estes sistemas devem ser melhorados e integrados em processos de gestão de desempenho mais robustos.

As transformações no mundo do trabalho reforçam essa necessidade. Mudanças como a digitalização, a expansão do trabalho remoto e os novos padrões de interação profissional tornam o desempenho mais difícil de observar e interpretar, exigindo sistemas avaliativos mais flexíveis e contínuos (Murphy & DeNisi, 2023, p. 144).

A crescente valorização das competências comportamentais também acompanha esta evolução. Em setores de serviços, por exemplo, a forma como os trabalhadores interagem com clientes e parceiros influencia diretamente a percepção de qualidade e a reputação institucional. Shury et al. (2010, como citados em Wilkinson et al., 2017, p. 140) mostram que competências como comunicação, postura, atitude e capacidade de lidar com clientes são consideradas essenciais pelos empregadores. Estas dimensões ampliam a importância de sistemas avaliativos capazes de captar capacidades técnicas e comportamentais subtis, bem como contributos técnicos e funcionais.

2.3. Modelos, métodos e critérios de avaliação do desempenho

Há vários métodos utilizados para avaliar desempenho, os quais variam dependendo dos objetivos organizacionais, dos tipos de competências avaliadas e do nível de estruturação desejado. Os modelos, em geral, podem ser agrupados em métodos tradicionais e contemporâneos, cada um com vantagens e limitações diferentes.

As abordagens tradicionais incluem escalas de classificação, listas de verificação, apreciações narrativas e relatórios escritos. Estes instrumentos são comumente utilizados pela sua simplicidade, mas costumam ser influenciados pela percepção subjetiva do avaliador, o que aumenta a probabilidade de erros, tal como observado por Javidmehr e Ebrahimpour (2015, p. 288).

As Escalas Comportamentais Ancoradas (BARS) e a técnica dos incidentes críticos são modelos mais recentes que procuram atenuar as limitações dos métodos tradicionais, aumentando a objetividade da avaliação ao basearem os julgamentos em comportamentos observáveis e em episódios reais de desempenho. Em conjunto, estes métodos reduzem a subjetividade e oferecem uma base mais sólida para apoiar decisões relacionadas com o desempenho, tal como descrito por Lunenburg (2012, p. 5).

Para além dos métodos, os critérios avaliativos também variam, abrangendo competências técnicas, comportamentais e funcionais. Armstrong (2015, p. 83) destaca que para que haja consistência e transparência, é importante definir critérios claros e alinhados com comportamentos observáveis. Modelos atuais integram ainda feedback contínuo, participação do colaborador na definição de metas e reuniões de acompanhamento regulares (Aguinis, 2014, p. 29). Estas são ações que objetivam corrigir limitações históricas e construir sistemas avaliativos mais fiáveis, justos, transparentes e que levem em consideração a complexidade do comportamento humano no local de trabalho.

2.4. Erros e vieses na avaliação de desempenho

Apesar dos avanços metodológicos, a avaliação de desempenho permanece sujeita a erros e vieses que distorcem julgamentos e comprometem a credibilidade do processo. Javidmehr e Ebrahimpour (2015, p. 288) descrevem erros que decorrem de distorções cognitivas, percepções subjetivas e influências contextuais.

Desse modo, os erros de avaliação podem ser compreendidos como distorções que comprometem a precisão, a justiça ou a utilidade do julgamento avaliativo. Os vieses, por sua vez, correspondem a tendências cognitivas, sociais ou contextuais que influenciam a

forma como o avaliador observa, interpreta e classifica o desempenho. Assim, o viés pode ser entendido como uma das causas ou mecanismos explicativos do erro, enquanto o erro representa a manifestação da distorção no processo avaliativo.

Além disso, fatores sociais e cognitivos também influenciam os julgamentos. Murphy e DeNisi (2023, pp. 144–147) demonstram que expectativas prévias, relações interpessoais, estereótipos e pressões organizacionais moldam a interpretação do desempenho, e que estes vieses persistem mesmo em sistemas estruturados ou acompanhados de formação, o que reforça a necessidade de abordagens avaliativas mais estruturadas, mais sólidas, criteriosas e transparentes.

2.4.1. Distorções clássicas na avaliação de avaliação

Javidmehr e Ebrahimpour (2015, p. 288) apresentam uma sistematização abrangente dos erros mais frequentes na avaliação de desempenho, isto é, distorções repetitivas que influenciam a forma como o desempenho é interpretado e classificado. A Tabela II apresenta os principais erros identificados pelos autores, com breves descrições que mostram o modo como cada um deles pode comprometer a precisão e a justiça do processo de avaliação.

TABELA II
ERROS E VIESES CLÁSSICOS NA AVALIAÇÃO DE DESEMPENHO

Distorção Avaliativa	Breve descrição
Efeito <i>Halo</i>	Uma característica positiva influencia outras dimensões da avaliação.
Efeito <i>Horn</i>	Uma característica negativa contamina a avaliação global.
Leniência	O avaliador atribui classificações sistematicamente mais altas do que o desempenho justifica.

Distorção Avaliativa	Breve descrição
Severidade	O avaliador atribui classificações sistematicamente mais baixas do que o desempenho justifica.
Tendência central	O avaliador evita extremos e concentra as classificações na zona média, reduzindo a diferenciação entre desempenhos.
Contraste	A avaliação é influenciada pela comparação com desempenhos de outros colaboradores, em vez de critérios objetivos.
Recência	Acontecimentos próximos da data da avaliação recebem um peso desproporcional, distorcendo a análise do período total.

Nota. Síntese elaborada pelo autor com base na sistematização de erros de avaliação discutida por Javidmehr e Ebrahimpour (2015).

A presença destes erros mostra como a avaliação de desempenho pode ser influenciada pelas percepções dos avaliadores, em contraste com o desempenho real dos colaboradores, o que compromete a equidade, a precisão e a justiça do processo avaliativo.

2.4.2. Vieses cognitivos e influências contextuais

Existem vieses cognitivos e dinâmicas contextuais que operam de forma menos visível do que os erros, mas que influenciam os julgamentos dos avaliadores. Murphy e DeNisi (2023, pp. 145–147) explicam que percepções iniciais, familiaridade, afinidade, expectativas prévias e estereótipos funcionam como filtros interpretativos que afetam a forma como o desempenho é percebido. Estes vieses não resultam de falhas técnicas dos instrumentos avaliativos, mas de mecanismos psicológicos automáticos que influenciam a atenção, a memória e o processamento de informação.

O contexto organizacional também desempenha um papel determinante. As pressões para gerenciar relações harmoniosas, normas implícitas de reciprocidade,

culturas organizacionais e políticas internas podem servir para desvirtuar julgamentos de forma sistemática. De acordo com Murphy e DeNisi (2023, pp. 146–147), essas influências persistem mesmo quando escalas estruturadas ou programas de treinamento para avaliadores são implementados, pois esses mecanismos cognitivos e sociais não são eliminados apenas por uma mudança técnica no sistema.

Assim, os vieses cognitivos e contextuais mostram que a avaliação de desempenho é um processo basicamente interpretativo, onde fatores psicológicos e sociais se entrelaçam com a análise do comportamento que se pode observar. A redução desses vieses exige abordagens que promovam consciência crítica, transparência e responsabilidade, sem deixar de reconhecer que a objetividade absoluta é difícil de alcançar em avaliações humanas.

2.5. Consequências dos erros na avaliação de desempenho

Os erros e vieses presentes nos sistemas de avaliação têm impactos significativos tanto para indivíduos como para organizações. Quando avaliações são percebidas como injustas, incoerentes ou subjetivas, ocorrem reduções na motivação, no envolvimento e na satisfação dos colaboradores. Wilkinson et al. (2017, p. 190) explicam que avaliações injustas enfraquecem a confiança na gestão e comprometem a credibilidade dos processos de gestão, afetando diretamente a moral e o desempenho das equipes.

No plano organizacional, DeNisi e Murphy (2017, p. 426) demonstram que a forma como a avaliação de desempenho é realizada influencia a produtividade, a equidade interna e o desenvolvimento do capital humano. No plano econômico, Cammeraat et al. (2021) reforçam que sistemas avaliativos claros e bem implementados estão associados a ganhos de produtividade e maior eficiência nas organizações. Os autores acrescentam

ainda que práticas avaliativas coerentes, meritocráticas e consistentes com o desenvolvimento do capital humano favorecem ambientes de trabalho mais unidos e reduzem desigualdades internas, reforçando a importância de sistemas de avaliação que promovam equidade, eficiência e sustentabilidade organizacional.

2.6. Inconsistências e limitações teóricas identificadas

Segundo Lameque et al. (2023, pp. 3-4), a literatura sobre avaliação de desempenho, no entanto, reflete fragmentação conceitual e inconsistências metodológicas, com diferenças nas definições de conceitos, explicações sobre o papel do avaliador e nas recomendações propostas pelos estudos. Estas diferenças tornam difícil comparar resultados, formular modelos e construir abordagens integradoras.

Parte da explicação para essa inconsistência pode ser atribuída à variedade de contextos organizacionais, métodos e finalidades para sistemas de avaliação. No entanto, evidenciam também a falta de um enquadramento suficientemente consolidado que facilite compreender de forma clara quais erros e vieses são mais recorrentes e que práticas são mais recomendadas para melhorar a qualidade das avaliações. Esta falta de convergência destaca a importância de uma síntese sólida e abrangente que reúna estudos recentes, identifique padrões e clarifique recomendações fundamentadas, contribuindo para o aperfeiçoamento das práticas avaliativas (Lameque et al., 2023, p. 3).

3. METODOLOGIA

A presente investigação consistiu numa Revisão Sistemática da Literatura (RSL), orientada pelas recomendações do PRISMA 2020 (Page et al., 2021). Este protocolo foi adotado por estabelecer critérios explícitos de identificação, triagem, elegibilidade e

inclusão dos estudos, conferindo maior transparência, rastreabilidade e consistência ao processo de seleção. O objetivo da revisão foi identificar os erros e vieses mais recorrentes nos sistemas de avaliação de desempenho e mapear as recomendações propostas pela literatura científica entre 2010 e 2025. A opção por uma abordagem sistemática justifica-se pela necessidade de organizar uma literatura fragmentada, reduzir escolhas arbitrárias na seleção dos estudos e construir uma base analítica coerente para a interpretação dos resultados.

As revisões sistemáticas permitem integrar resultados de múltiplos estudos (Page et al., 2021), proporcionando uma visão consolidada do estado da arte, especialmente em temas fragmentados ou em evolução conceptual, como é o caso da avaliação de desempenho. Briner e Denyer (2012) salientam que as RSL introduzem processos padronizados para recolha, análise e seleção dos estudos, contribuindo para um corpo de conhecimento mais robusto e comparável. Munn et al. (2018, p. 5) mencionam que a utilização de critérios explícitos de elegibilidade reduz o risco de enviesamento e fortalece a validade das conclusões. Gough et al. (2012, pp. 2-3), corroboram a relevância deste método, os quais destacam o papel das revisões sistemáticas na identificação de padrões teóricos, inconsistências e direções para investigações futuras.

A literatura confirma ainda a fiabilidade das bases de dados utilizadas neste estudo. A *EBSCO Discovery Service* é amplamente reconhecida pela sua abrangência, congregando conteúdos provenientes de bases consolidadas (EconLit, APA PsycInfo, Business Source, entre outras) e permitindo pesquisas simultâneas em milhares de periódicos académicos com revisão por pares (Alves & Mariano, 2018, p. 260). Paralelamente, a *Scopus* e a *Web of Science* (WoS) destacam-se pela sua qualidade, rigor de indexação e cobertura multidisciplinar, sendo consideradas referências essenciais em

estudos de natureza bibliométrica e sistemática (Cobo et al., 2011; Dangelico, 2016). A utilização combinada destas três plataformas garante amplitude, diversidade e fiabilidade na identificação dos estudos relevantes.

3.1. *Estratégia de pesquisa*

A estratégia de pesquisa foi delimitada objetivando responder diretamente à questão de investigação. Para assegurar rigor e exaustividade, a pesquisa foi realizada em três bases de dados de elevada credibilidade: EBSCO, *Scopus* e *Web of Science*.

As buscas seguiram uma estrutura progressiva, organizada em três blocos de descritores:

1. Bloco 1 – Tema principal

“performance appraisal” OR “employee performance evaluation”

2. Bloco 2 – Erros e vieses avaliativos

“cognitive bias” OR “rating errors” OR “evaluation bias”

3. Bloco 3 – Recomendações e práticas de melhoria

“best practices” OR “recommendations” OR “training of raters” OR “feedback process”

A pesquisa sobre o tema principal foi realizada apenas nos títulos, enquanto os demais termos foram pesquisados em todos os campos, permitindo refinar os resultados e garantir que os estudos selecionados abordassem diretamente os temas de interesse.

3.2. *Justificação da baliza temporal*

O intervalo temporal entre 2010 e 2025 foi definido com base em critérios teóricos e metodológicos. A literatura publicada até o início da década de 2010 foi amplamente

consolidada por autores como Kondrasuk (2011, p. 57), que sintetizou 76 problemas clássicos dos sistemas avaliativos, representando um marco na sistematização dos erros historicamente identificados.

A partir de 2010, observa-se uma transformação significativa na literatura sobre avaliação de desempenho, marcada pelo crescimento de abordagens de feedback contínuo, pela intensificação dos debates sobre justiça e vieses cognitivos e pela influência crescente da digitalização e do trabalho híbrido na observação do desempenho. Autores como DeNisi e Murphy (2017, pp. 421-423), Aguinis (2014, p. 29) e Murphy e DeNisi (2023, p. 144-146) destacam que este período corresponde a uma fase conceptual distinta, caracterizada por renovação dos modelos, revisão das práticas tradicionais e emergência de novos desafios avaliativos. Assim, a delimitação temporal 2010–2025 permite captar estas transformações e analisar as recomendações propostas pela literatura contemporânea.

3.3. Justificação dos idiomas incluídos

Foram incluídos artigos em inglês e português. O inglês é o idioma predominante na produção científica internacional em gestão e psicologia organizacional, garantindo acesso às contribuições mais influentes e atualizadas. A inclusão do português justifica-se pela relevância da literatura produzida em contextos lusófonos, especialmente Portugal e Brasil, onde se desenvolvem investigações importantes sobre práticas avaliativas, percepções de justiça e desafios organizacionais. Esta escolha assegura abrangência conceptual e representatividade sem comprometer a uniformidade metodológica.

3.4. Resultados da pesquisa nas bases de dados

A pesquisa nas bases de dados resultou na identificação inicial de 184 registros, posteriormente filtrados conforme os critérios temporais, documentais e de revisão por pares definidos para esta investigação. A Tabela III sintetiza os resultados obtidos em cada base de dados.

TABELA III
RESULTADOS DA PESQUISA NAS BASES DE DADOS

Etapa do processo	EBSCO	Scopus	Web of Science	Total
Registos identificados	157	24	3	184
Removidos por serem anteriores a 2010	88	4	1	93
Duplicados removidos	32	1	1	34
Não revistos por pares	4	0	0	4
Excluídos por tipo de documento	0	5	1	6
Registos restantes para triagem	33	14	0	47
Artigos incluídos no corpus final	25	12	0	37

Nota. Elaboração própria com base nos resultados da pesquisa realizada nas bases de dados EBSCO, Scopus e Web of Science. A exclusão por tipo de documento incluiu artigos de conferência, livros e capítulos de livro incompatíveis com os critérios de elegibilidade definidos. A inclusão final resultou da triagem, da recuperação dos textos integrais e da leitura completa dos estudos potencialmente elegíveis.

3.5. Critérios de elegibilidade

Critérios de inclusão

- Estudos que analisam erros, vieses cognitivos, falhas de processo ou limitações estruturais da avaliação de desempenho
- Estudos que apresentam recomendações práticas, melhorias, modelos, técnicas ou formação para qualificar sistemas avaliativos

- Estudos empíricos, teóricos ou de revisão publicados em revistas indexadas e com revisão por pares
- Publicação entre 2010 e 2025

Critérios de exclusão

- Estudos que analisam apenas desempenho organizacional, motivação, produtividade, envolvimento ou gestão de desempenho sem foco em avaliação de desempenho
- Estudos centrados em ferramentas específicas (competências, escalas motivacionais, sistemas tecnológicos) sem ligação direta ao problema dos erros e vieses
- Modelos de decisão multicritério, análises de eficiência e estudos de implementação cultural sem contributo conceptual para erros ou recomendações
- Documentos não considerados artigos científicos (artigos de conferência, livros e capítulos de livros, relatórios, teses)

3.6. Processo de seleção dos estudos

A pesquisa realizada nas bases de dados EBSCO, *Scopus* e *Web of Science* identificou inicialmente 184 registos. Numa primeira fase, foram removidos 93 estudos publicados antes de 2010, 34 registos duplicados, quatro artigos não revistos por pares e seis documentos incompatíveis com os critérios definidos (dois artigos de conferência, dois livros e dois capítulos de livro).

Após estas primeiras exclusões, permaneceram 47 estudos potencialmente elegíveis para triagem. Na fase de triagem por título e resumo, quatro artigos foram excluídos por não apresentarem relação suficientemente direta com os objetivos desta investigação, incidindo predominantemente sobre modelos multicritério de apoio à

decisão, gestão do desempenho em sentido amplo ou abordagens sem contributo conceptual relevante para erros, vieses ou falhas estruturais da avaliação de desempenho.

Assim, 43 estudos foram considerados potencialmente relevantes e submetidos à fase de recuperação do texto integral. Destes, dois artigos não puderam ser recuperados integralmente, apesar de tentativas de contacto com os autores e pesquisa em diferentes bases de dados e repositórios académicos.

Na fase de elegibilidade, 41 artigos foram analisados integralmente. Após leitura completa, quatro estudos foram excluídos por não satisfazerem os critérios de elegibilidade definidos, nomeadamente por apresentarem foco predominantemente centrado em qualidade geral da avaliação, aceitação do sistema avaliativo, gestão de talentos ou fatores motivacionais sem relação suficientemente direta com erros, vieses cognitivos ou falhas estruturais da avaliação de desempenho.

A amostra final foi composta por 37 artigos, que constituem a base empírica e conceptual desta revisão sistemática da literatura.

3.7. *Fluxograma PRISMA*

O processo de seleção dos estudos foi sintetizado conforme o PRISMA 2020, indicando as fases de identificação, triagem, elegibilidade e inclusão. A Figura 1 apresenta o fluxograma PRISMA que sintetiza o processo de seleção.

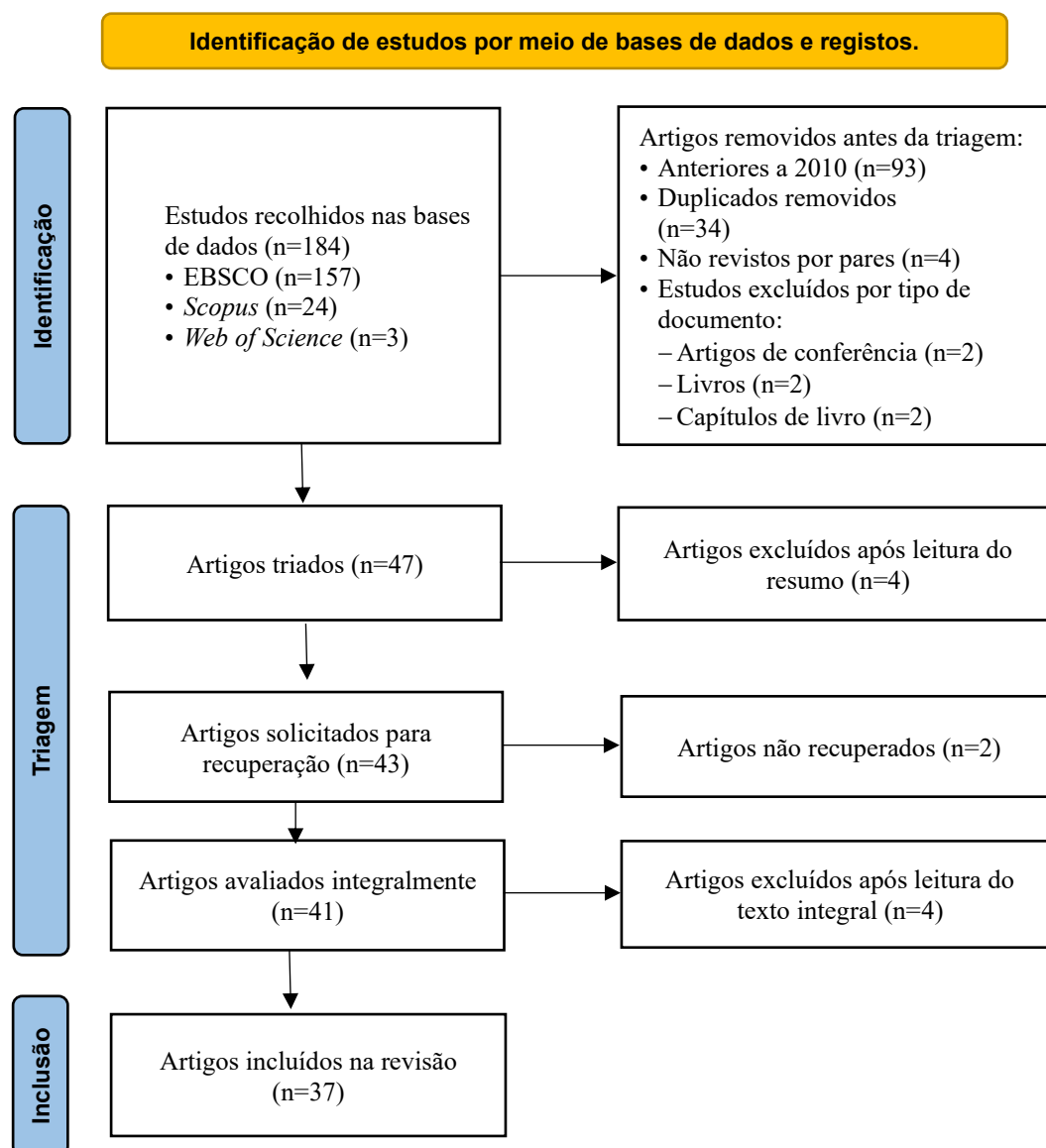


FIGURA 1 - Fluxograma PRISMA para a Apresentação do Processo de Seleção dos Estudos

Nota. Elaboração própria com base no processo de identificação, triagem, elegibilidade e inclusão dos estudos, conforme a estrutura PRISMA 2020 descrita na metodologia.

3.8. Extração e organização dos dados

Para cada artigo incluído, foram extraídas informações relativas a:

- autor(es) e ano;
- país de origem ou contexto do estudo;
- objetivos do artigo;

- tipo de abordagem (conceptual, empírica, revisão etc.);
- erros, vieses ou limitações identificadas;
- recomendações, práticas ou sugestões apresentadas;
- conclusões principais.

Estas informações foram organizadas numa matriz de extração e análise, desenvolvida especificamente para esta investigação, e serviram de base para a síntese dos resultados.

A matriz foi estruturada de modo a responder não apenas à questão principal da investigação, mas também às subquestões analíticas definidas para esta revisão. Assim, os dados extraídos permitiram identificar erros e vieses recorrentes, falhas estruturais associadas, recomendações propostas pela literatura e os diferentes níveis de análise em que essas soluções se concentram.

3.9. *Estratégia de análise*

A análise dos dados seguiu uma abordagem qualitativa e interpretativa, orientada para a síntese temática dos estudos incluídos. Os artigos foram analisados com base em categorias previamente definidas e refinadas durante o processo de leitura integral, permitindo:

- identificar os erros, vieses e limitações estruturais mais recorrentes nos sistemas de avaliação de desempenho;
- compreender os fatores contextuais, processuais e organizacionais associados às distorções avaliativas;
- classificar as recomendações e práticas de mitigação segundo o tipo de mecanismo indicado pela literatura;

- categorizar as soluções identificadas segundo diferentes níveis de análise (individual, interpessoal, organizacional e estratégico);
- identificar convergências, divergências, lacunas e tendências emergentes na literatura entre 2010 e 2025.

Esta estratégia permitiu não apenas descrever os problemas identificados na literatura, mas também analisar criticamente a forma como os estudos recentes procuram responder às limitações históricas dos sistemas de avaliação de desempenho.

4. ANÁLISE DOS RESULTADOS

A análise dos 37 artigos incluídos na revisão sistemática permitiu identificar padrões recorrentes sobre erros, vieses, falhas estruturais e recomendações nos sistemas de avaliação de desempenho.

Importa ressaltar que as contagens apresentadas não representam prevalência estatística universal dos problemas nas organizações, mas sim recorrência temática das categorias codificadas na matriz de extração e análise. Assim, um problema identificado em maior número de artigos deve ser interpretado como mais recorrente no corpus analisado, e não necessariamente como mais frequente em todos os sistemas de avaliação de desempenho.

Também é importante destacar que os artigos incluídos não apresentam o mesmo tipo de evidência. Alguns estudos testam diretamente determinados vieses ou mecanismos de mitigação, enquanto outros discutem esses fenômenos de forma conceptual, documental ou a partir de percepções dos participantes. Por isso, a análise distingue, sempre que necessário, entre evidência empírica direta, evidência empírica percebida ou indireta e apoio conceptual.

4.1. Caracterização geral dos estudos incluídos

A distribuição temporal dos artigos, apresentada na Figura 2, evidencia a cobertura de todo o período definido para a revisão, entre 2010 e 2025.

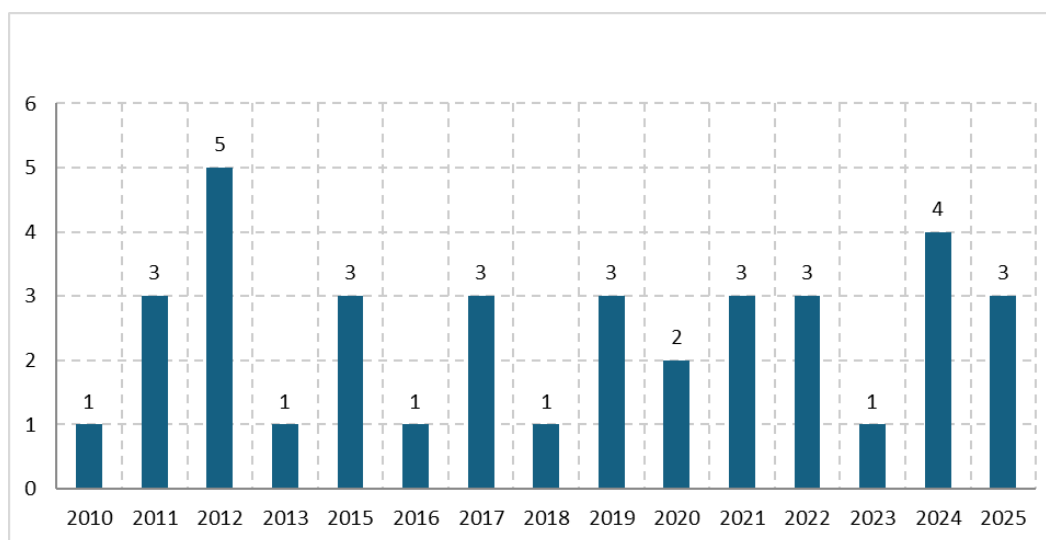


FIGURA 2 - Gráfico - Distribuição temporal dos artigos

Nota. Elaboração própria com base na matriz de extração e análise dos 37 artigos incluídos na revisão sistemática.

Observa-se, portanto, que a avaliação de desempenho permanece como tema recorrente na literatura entre 2010 e 2025. A concentração observada em 2012 e 2024 não altera a cobertura do período analisado, mas indica momentos de maior produção dentro do corpus selecionado. Essa distribuição também demonstra que o tema continua atual, especialmente diante das discussões recentes sobre justiça percebida, feedback contínuo, cultura organizacional, digitalização e limitações dos modelos tradicionais de avaliação.

Para complementar a caracterização do corpus, a Tabela IV apresenta a natureza metodológica dos estudos incluídos.

TABELA IV
NATUREZA DOS ESTUDOS INCLUÍDOS

Natureza do estudo	Quantidade de artigos	Interpretação analítica
Empírico	24	Estudos com dados próprios, incluindo inquéritos, experiências, entrevistas, estudos de caso, análise documental ou dados organizacionais.
Teórico ou revisão	10	Estudos usados para enquadramento, sistematização de conceitos, modelos explicativos e recomendações.
Documental ou analítico	2	Estudos baseados em análise institucional, normativa ou documental.
Misto	1	Estudo que combina revisão de literatura, estudo de caso organizacional, entrevistas orais e análise documental.

Nota. Elaboração própria com base na classificação metodológica registada na matriz de extração e análise dos 37 artigos incluídos na revisão sistemática.

Quanto à natureza dos estudos, observa-se predominância de artigos empíricos, correspondentes a 24 dos 37 trabalhos incluídos. Esse resultado é relevante porque indica que a revisão não se apoia apenas em textos conceptuais, mas também em investigações com dados próprios, incluindo inquéritos, experiências, entrevistas, estudos de caso, análise documental, dados organizacionais e validação de instrumentos. Ao mesmo tempo, os artigos teóricos e de revisão cumprem função importante, pois ajudam a organizar conceitos, explicar categorias e interpretar recomendações presentes na literatura.

Para complementar a caracterização do corpus, foi analisada a abordagem metodológica predominante dos estudos incluídos. A Figura 3 apresenta essa distribuição, permitindo observar a diversidade metodológica da revisão.

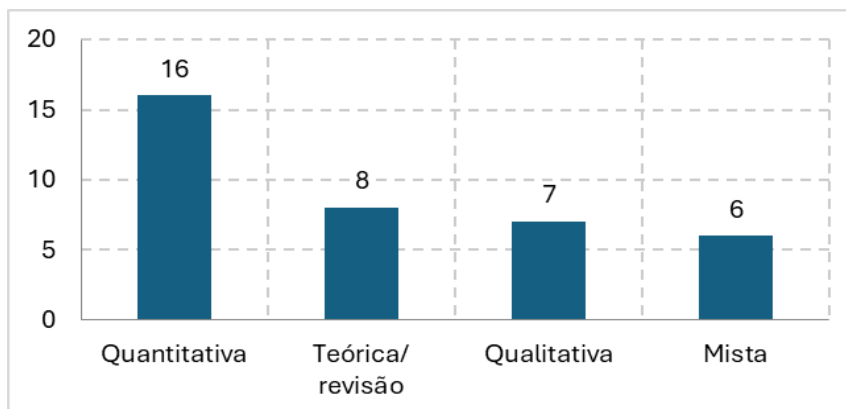


FIGURA 3 - Gráfico - Abordagem Metodológica Predominante

Nota. Elaboração própria com base na classificação da abordagem metodológica predominante registrada na matriz de extração e análise dos 37 artigos incluídos na revisão sistemática.

A diversidade metodológica exige cautela na leitura dos achados. Estudos experimentais, como Belle et al. (2017), Onal e Kumkale (2022) e Kim e Campbell (2025), oferecem evidência mais direta sobre vieses específicos. Estudos qualitativos e contextuais, como Haque (2012), Liu e Dong (2012), Yahiaoui et al. (2021), Makhubela e Malatji (2024) e Tarigan et al. (2025), ajudam a compreender fatores institucionais, culturais, processuais e relacionais associados à avaliação de desempenho. Enquanto estudos conceituais ou de revisão, como Kondrasuk (2011), Kromrei (2015), Iqbal et al. (2015), DeNisi e Murphy (2017), Larsen (2019) e Park (2017), contribuem para sistematizar categorias, interpretar problemas e identificar recomendações, mas não devem ser tratados como testes empíricos de prevalência dos vieses.

Dessa forma, a análise dos resultados considera não apenas a frequência com que determinada categoria aparece na matriz, mas também o tipo de evidência oferecido pelos estudos em que ela foi identificada. Essa precaução evita que menções conceituais sejam interpretadas como achados empíricos e preserva a coerência metodológica da revisão.

4.2. Problemas, erros e vieses mais recorrentes

A Tabela V sintetiza os problemas, erros e vieses mais recorrentes na literatura analisada.

TABELA V
CATEGORIAS ANALÍTICAS MAIS RECORRENTES ASSOCIADAS À AVALIAÇÃO DE DESEMPENHO

Categoria	Natureza da categoria	Quantidade de artigos	Interpretação analítica
Falhas de desenho e implementação	Problema sistémico	29	Explica por que a avaliação pode falhar antes mesmo do julgamento individual
Subjetividade do avaliador	Viés geral	24	Mostra que julgamentos avaliativos ficam mais vulneráveis quando há critérios vagos, baixa padronização e pouca rastreabilidade
Feedback insuficiente	Falha de processo	20	Mostra a fragilidade da avaliação quando ela não gera orientação para melhoria
Metas e critérios pouco claros	Falha estrutural	20	Aumenta a margem para interpretações subjetivas e distorções avaliativas
Contexto cultural e institucional	Viés contextual	19	Mostra que os sistemas não operam de forma neutra em relação ao contexto
Formação e calibração de avaliadores insuficientes	Falha de capacidade avaliativa	18	Aparece como necessidade recorrente para reduzir inconsistência e subjetividade
Leniência e severidade	Viés clássico	14	Expressa dificuldade de diferenciar adequadamente níveis de desempenho

Categoria	Natureza da categoria	Quantidade de artigos	Interpretação analítica
Justiça percebida e injustiça	Reação ao sistema	14	Relaciona erros e falhas processuais à aceitação, confiança e legitimidade do sistema
Efeito <i>halo</i> e efeito <i>horn</i>	Viés clássico	12	Mostra contaminação entre dimensões de desempenho
Favoritismo e política organizacional	Viés político-relacional	12	Indica que a avaliação também é afetada por relações de poder e interesses
Recência e memória	Viés temporal	10	Evidencia o risco de depender da memória no fim do ciclo avaliativo
Múltiplas fontes e 360 graus	Mecanismo de mitigação	9	Surge como solução possível, mas dependente de desenho e propósito claros
Tendência central	Viés clássico	8	Revela baixa capacidade discriminatória, evitação de conflito ou insegurança
Tecnologia e digitalização	Mecanismo de suporte	2	Apoia registo e rastreabilidade, mas não constitui solução autónoma

Nota. Elaboração própria com base na matriz de extração e análise dos 37 artigos incluídos na revisão sistemática. As contagens indicam recorrência temática das categorias codificadas, não prevalência estatística universal dos problemas nos sistemas de avaliação de desempenho.

A Tabela V mostra que os problemas mais recorrentes não estão concentrados apenas nos vieses clássicos do avaliador. A categoria mais frequente foi a de falhas de desenho e implementação, presente em 29 artigos, seguida pela subjetividade do avaliador, identificada em 24 artigos. Observa-se, então, que parte expressiva das fragilidades da avaliação de desempenho está associada à arquitetura do sistema, à clareza

dos critérios, à qualidade da comunicação, à preparação dos avaliadores e ao uso posterior dos resultados.

Essa leitura é coerente com estudos que tratam a avaliação de desempenho como um sistema organizacional e não apenas como um ato isolado de julgamento. Kondrasuk (2011) e Iqbal et al. (2015) ajudam a compreender essa dimensão ao discutirem a importância do desenho do sistema, da finalidade da avaliação, dos critérios utilizados, dos intervenientes e das consequências associadas ao processo. Em contextos empíricos específicos, Agbola et al. (2011), Haque (2012), Liu e Dong (2012), McClendon et al. (2020), Makhubela e Malatji (2024) e Tarigan et al. (2025) também mostram que a avaliação perde credibilidade quando é aplicada de forma irregular, pouco transparente ou sem ligação clara com critérios, feedback, desenvolvimento, reconhecimento e melhoria organizacional.

Os vieses clássicos continuam relevantes, mas aparecem articulados a condições que os favorecem. Leniência, severidade, efeito *halo*, efeito *horn*, recência e tendência central não devem ser interpretados apenas como falhas cognitivas individuais, pois tendem a ser agravados por critérios ambíguos, ausência de registros, baixa calibração e fraca responsabilização. Belle et al. (2017), Merkel et al. (2021), Onal e Kumkale (2022), Odetunde (2022) e Kim e Campbell (2025) discorrem sobre distorções de julgamento, incluindo efeito *halo*, enviesamentos de avaliação, recência, memória, otimismo, positividade e correspondência de desempenho (*performance matching*). Kumari (2012), Kromrei (2015), Larsen (2019) e Tshukudu (2020), por sua vez, ajudam a enquadrar esses erros como problemas recorrentes da literatura e de contextos aplicados de avaliação de desempenho.

A presença de categorias como contexto cultural, justiça percebida, favoritismo, política organizacional e fatores ambientais amplia a compreensão dos problemas. A avaliação de desempenho não ocorre em ambiente neutro, pois é influenciada por relações hierárquicas, normas culturais, redes informais, interesses organizacionais, percepções de justiça e circunstâncias específicas do contexto em que é realizada. Essa dimensão aparece em estudos como os de Appelbaum et al. (2011), Bizzi (2018), Prati et al. (2019), Sanner et al. (2022), Yahiaoui et al. (2021), Daud et al. (2023), Micacchi et al. (2024) e Tarigan et al. (2025), nos quais a avaliação é analisada como prática social, relacional, institucional e contextual.

A Figura 4 complementa a leitura da Tabela V ao representar a coocorrência entre as categorias codificadas. No mapa, os nós indicam as categorias, o tamanho reflete o peso da categoria, as ligações demonstram presença conjunta no mesmo artigo e as cores representam agrupamentos temáticos gerados pelo software VOSviewer.

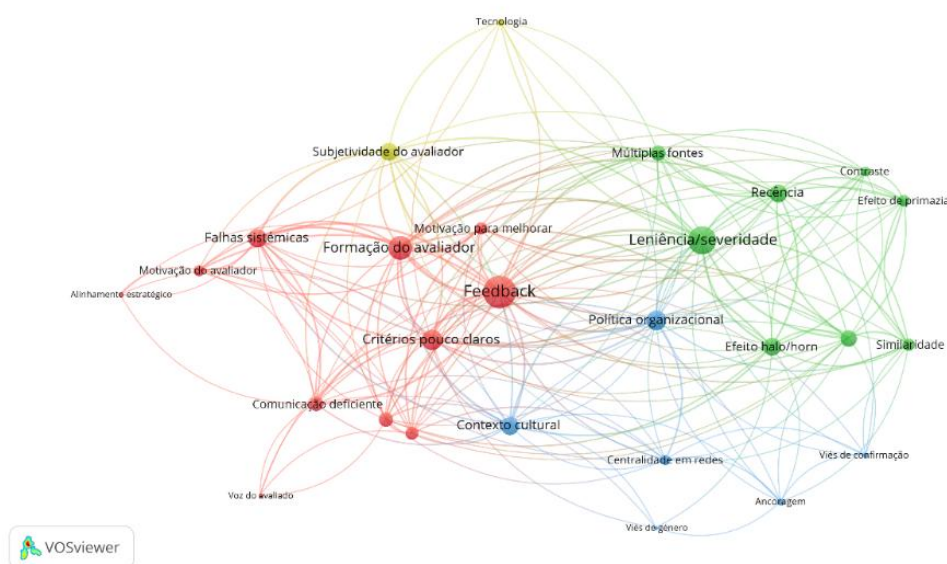


FIGURA 4 - Mapa de coocorrência das categorias codificadas

Nota. Elaboração própria com base nas categorias codificadas nos 37 artigos incluídos na revisão sistemática, associadas a problemas, erros, vieses e mecanismos de mitigação em avaliação de desempenho. A visualização foi produzida no software VOSviewer e deve ser interpretada como mapa exploratório de coocorrência, não como análise bibliométrica em sentido estrito. Os rótulos foram abreviados para melhorar a legibilidade visual.

A Figura 4 reforça que os problemas da avaliação de desempenho não aparecem de forma isolada. As categorias aproximam-se em três núcleos principais: um sistêmico e processual, associado ao desenho do sistema, aos critérios, ao feedback e à formação dos avaliadores; outro ligado aos vieses clássicos e cognitivos, como efeito *halo*, efeito *horn*, leniência, severidade, recência e tendência central; e um terceiro associado a fatores contextuais, culturais, relacionais e político-organizacionais. Dessa forma, o mapa confirma a leitura já indicada na Tabela V: os erros de avaliação não decorrem apenas do julgamento individual do avaliador, mas da interação entre critérios, processo, cultura, documentação, feedback e condições institucionais.

Diante disso, os erros e vieses na avaliação de desempenho não devem ser lidos de forma separada. Mesmo observando que os vieses individuais continuam relevantes, os resultados indicam que eles se articulam com falhas estruturais, processuais e contextuais. Sendo assim, para que a avaliação melhore, é necessário que haja respostas integradas, capazes de atuar sobre critérios, feedback, formação, calibração, documentação, transparência e cultura organizacional.

4.3. *Articulação entre problemas e mecanismos de mitigação*

Os principais problemas identificados na matriz foram articulados com os mecanismos de mitigação sugeridos pela literatura, conforme apresentado na Tabela VI. Essa articulação não pretende estabelecer soluções universais, mas demonstrar como os problemas recorrentes se relacionam com respostas possíveis nos níveis individual, interpessoal, organizacional e estratégico.

TABELA VI

QUADRO ANALÍTICO ENTRE PROBLEMAS RECORRENTES E MECANISMOS DE MITIGAÇÃO

Problema identificado	Natureza do problema	Recorrência	Mecanismo de mitigação	Interpretação analítica
Falhas de desenho e implementação	Problema sistêmico	29	Rever a arquitetura do sistema, a política escrita, os papéis dos intervenientes, a periodicidade e o uso dos resultados	Mostra que a avaliação pode falhar antes mesmo do julgamento individual, quando o sistema é mal concebido ou mal executado.
Subjetividade do avaliador	Viés geral	24	Padronizar critérios, usar evidências observáveis, formar avaliadores e criar mecanismos de revisão	Liga os erros clássicos a critérios vagos, baixa padronização e fraca governança
Feedback insuficiente	Falha de processo	20	Institucionalizar feedback contínuo, específico e orientado para a melhoria	Indica que a avaliação perde valor quando não orienta a melhoria futura
Metas e critérios pouco claros	Falha estrutural	20	Definir metas, indicadores, rubricas e expectativas observáveis	Reduz a margem para interpretações subjetivas e torna o processo mais compreensível
Contexto cultural e institucional	Viés contextual	19	Adaptar o sistema ao contexto organizacional, cultural e institucional	Mostra que práticas avaliativas não operam de forma neutra e podem ser reinterpretadas localmente
Formação e calibração insuficientes	Falha de capacidade avaliativa	18	Formar avaliadores, alinhar padrões de julgamento e calibrar classificações	Reforça que a consistência da avaliação depende da preparação de quem avalia

Problema identificado	Natureza do problema	Recorrência	Mecanismo de mitigação	Interpretação analítica
Leniência e severidade	Viés clássico	14	Calibrar classificações, exigir justificção das classificações e rever discrepâncias entre avaliadores	Expressa a dificuldade de diferenciar adequadamente níveis de desempenho
Justiça percebida e injustiça	Reação ao sistema	14	Aumentar transparência, participação, explicação dos critérios e coerência no uso dos resultados	Relaciona erros e falhas processuais à aceitação e legitimidade do sistema
Efeito <i>halo</i> e efeito <i>horn</i>	Viés clássico	12	Separar dimensões de desempenho, usar âncoras comportamentais e evidências específicas	Mostra a contaminação entre dimensões avaliativas distintas
Favoritismo e política organizacional	Viés político-relacional	12	Reforçar auditoria, múltiplas fontes, critérios observáveis e rastreabilidade das decisões	Indica que a avaliação também é afetada por relações de poder e interesses
Recência e memória	Viés temporal	10	Registrar evidências ao longo do ciclo avaliativo	Reduz a dependência da memória no momento final da avaliação
Múltiplas fontes e 360 graus	Mecanismo de mitigação	9	Usar fontes complementares com finalidade clara, confidencialidade e critérios definidos	Pode enriquecer a avaliação, desde que não seja aplicado de forma mecânica
Tendência central	Viés clássico	8	Melhorar escalas, calibrar padrões e exigir diferenciação justificada	Revela baixa capacidade discriminatória ou evitação de conflito avaliativo

Problema identificado	Natureza do problema	Recorrência	Mecanismo de mitigação	Interpretação analítica
Tecnologia e digitalização	Suporte ao processo	2	Apoiar registo, rastreabilidade, acompanhamento e padronização	Funciona como infraestrutura de suporte, mas não substitui critérios claros, justiça e qualidade do feedback

Nota. Elaboração própria com base na matriz de extração e análise dos 37 artigos incluídos na revisão sistemática. A recorrência indica o número de artigos em que cada categoria foi codificada, não uma medida de prevalência estatística universal dos problemas nos sistemas de avaliação de desempenho.

A Tabela VI permite observar que os erros e vieses de avaliação não devem ser interpretados de forma individual. Falhas de desenho e implementação, subjetividade do avaliador, feedback insuficiente e critérios pouco claros aparecem como categorias centrais porque condicionam a própria qualidade do julgamento avaliativo. Nessa perspectiva, os vieses clássicos, como leniência, severidade, efeito *halo*, efeito *horn*, recência e tendência central, não aparecem apenas de limitações cognitivas individuais, mas também de sistemas que oferecem pouca orientação, baixa rastreabilidade, reduzida calibração e fraca responsabilização.

Essa leitura encontra apoio em estudos que tratam a avaliação como sistema organizacional, como Kondrasuk (2011), Iqbal et al. (2015), McClendon et al. (2020), Micacchi et al. (2024) e Tarigan et al. (2025). Ao mesmo tempo, estudos experimentais e aplicados, como Belle et al. (2017), Merkel et al. (2021), Onal e Kumkale (2022), Odetunde (2022) e Kim e Campbell (2025), reforçam que o julgamento avaliativo permanece vulnerável a distorções cognitivas e temporais. Assim, a matriz permite afirmar que a mitigação dos erros exige atuação simultânea sobre o avaliador, os critérios, o processo, os registros, o feedback e a cultura organizacional.

Dessa forma, a formação dos avaliadores é necessária, mas insuficiente quando não existe clareza sobre o que deve ser avaliado, quando os resultados não são explicados, quando as evidências não são documentadas e quando a organização não dispõe de mecanismos de calibração ou revisão. Do mesmo modo, múltiplas fontes, tecnologia e feedback contínuo podem apoiar a melhoria do processo avaliativo, desde que estejam inseridos num sistema coerente, com finalidade clara, critérios partilhados e uso transparente dos resultados. O quadro articula, portanto, o diagnóstico dos problemas com os mecanismos de mitigação identificados na literatura.

4.4. *Recomendações recorrentes e níveis de análise*

As recomendações mais recorrentes identificadas na matriz de extração e análise concentram-se em mecanismos de clarificação, feedback, justiça, participação, treino e calibração, conforme sistematizado na Figura 5.

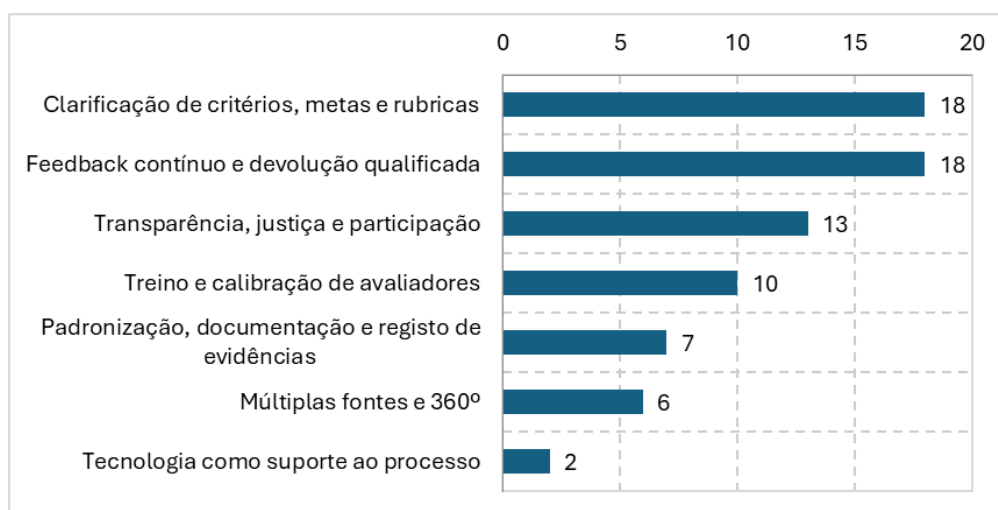


FIGURA 5 - Gráfico - Recomendações recorrentes identificadas na matriz de análise

Nota. Elaboração própria com base na matriz de extração e análise dos 37 artigos incluídos na revisão sistemática. A figura apresenta o número de artigos em que cada recomendação foi codificada; as categorias representam mecanismos recorrentes de mitigação dos problemas, erros e vieses identificados nos sistemas de avaliação de desempenho.

Como demonstra a Figura 5, as recomendações mais recorrentes concentram-se na clarificação de critérios, metas e rubricas e no fortalecimento do feedback contínuo e qualificado, ambas identificadas em 18 artigos. Esse padrão sugere que a literatura analisada não restringe a mitigação dos erros e vieses à correção do julgamento individual do avaliador, mas a associa ao aperfeiçoamento das condições processuais e organizacionais da avaliação. Essa leitura é coerente com McClendon et al. (2020), que evidenciam a relevância da execução do sistema pelo supervisor, e com Micacchi et al. (2024), ao relacionarem desenho da avaliação, feedback, calibração, justiça percebida e envolvimento dos trabalhadores.

A Tabela VII organiza as recomendações em quatro níveis de análise, evidenciando que a mitigação dos erros e vieses exige medidas combinadas.

TABELA VII
SÍNTESE DOS NÍVEIS DE ANÁLISE DAS RECOMENDAÇÕES

Nível de análise	Foco principal	Exemplos de mecanismos
Individual	Capacidade, motivação e consciência do avaliador	Formação, treino de vieses, treino de observação, preparação para avaliação.
Interpessoal	Relação avaliador-avaliado e qualidade da comunicação	Feedback contínuo, reuniões de acompanhamento, participação em metas, confiança.
Organizacional	Desenho, governança e consistência do sistema	Critérios claros, rubricas, documentação, calibração, transparência, justiça.
Estratégico	Integração da avaliação com objetivos e desempenho organizacional	Alinhamento estratégico, cultura de desempenho, tecnologia, ligação com desenvolvimento e decisões de GRH.

Nota. Elaboração própria com base na matriz de extração e análise dos 37 artigos incluídos na revisão sistemática. Os níveis foram organizados analiticamente a partir das recomendações codificadas.

A leitura conjunta da Figura 5 e da Tabela VII demonstra que a formação dos avaliadores é importante, mas não suficiente. No nível individual, estudos como Belle et al. (2017), Rosales Sánchez et al. (2019), Onal e Kumkale (2022) e Kim e Campbell (2025) indicam a necessidade de preparar os avaliadores para reconhecer distorções cognitivas, aplicar critérios e utilizar evidências de forma mais consistente. Contudo, essa preparação só produz melhores resultados quando articulada a critérios claros, documentação, calibração e mecanismos de revisão.

No nível interpessoal, destacam-se o feedback, a participação do avaliado e a qualidade da comunicação. Estudos como Kromrei (2015), Park (2017), Selvarajan e Cloninger (2012), Rivera et al. (2024), Micacchi et al. (2024) e Shim et al. (2025) mostram que a avaliação tende a ganhar legitimidade quando o avaliado compreende os critérios, participa do processo e recebe devoluções úteis, frequentes e contextualizadas para melhoria. Assim, a avaliação deixa de ser apenas classificação final e passa a funcionar também como instrumento de desenvolvimento.

Estudos de caso e análises aplicadas também reforçam essa leitura. Singh et al. (2010), Honarmand Shahzileh e Aghajan (2015) e Dohare e Singh (2021) mostram que a eficácia da avaliação depende da clareza dos critérios, da comunicação do processo, da formação dos avaliadores, do feedback ao avaliado e da ligação transparente entre avaliação, desenvolvimento, recompensas e decisões de gestão. Javidmehr e Ebrahimpour (2015), por sua vez, ajudam a sistematizar erros recorrentes de julgamento, como efeito de *halo*, efeito de *horn*, leniência, severidade, recência, tendência central, contraste e viés pessoal.

No nível organizacional e estratégico, concentram-se as medidas que dão sustentação ao sistema, como critérios observáveis, rubricas, registo de evidências, transparência, justiça, calibração, cultura orientada ao desempenho e alinhamento com decisões de gestão de recursos humanos. Essa dimensão aparece em estudos como Kondrasuk (2011), Haque (2012), Liu e Dong (2012), Rolle e Klingner (2012), Ayers (2013), Iqbal et al. (2015), Aro-Gordon (2016), McClendon et al. (2020), Botelho (2024), Makhubela e Malatji (2024) e Tarigan et al. (2025).

Dessa forma, os resultados indicam que para a diminuição dos erros e vieses deve haver coerência entre pessoas, processos, instrumentos e finalidade institucional.

4.5. *Síntese analítica dos resultados*

A análise dos 37 artigos permite concluir que os erros, vieses e falhas da avaliação de desempenho não se concentram apenas no momento do julgamento individual. Embora vieses clássicos, como efeito *halo*, efeito *horn*, leniência, severidade, recência e tendência central, permaneçam relevantes, os resultados mostram que esses problemas aparecem frequentemente associados a falhas de desenho, critérios pouco claros, feedback insuficiente, baixa documentação, reduzida calibração e condições culturais ou institucionais que afetam a forma como a avaliação é realizada.

A principal contribuição analítica está na articulação entre problemas e mecanismos de mitigação. A Tabela V identifica as categorias mais recorrentes, a Figura 4 evidencia as relações entre elas, a Tabela VI organiza a correspondência entre problemas e respostas possíveis, enquanto a Figura 5 e a Tabela VII sintetizam as recomendações por recorrência e por nível de análise. Em conjunto, esses elementos indicam que a melhoria dos sistemas avaliativos depende da coerência entre critérios,

comunicação, feedback, formação, documentação, justiça percebida, tecnologia de suporte e alinhamento estratégico.

Assim, os resultados reforçam que a mitigação dos erros e vieses exige uma abordagem integrada. A formação dos avaliadores continua sendo necessária, especialmente diante das distorções de julgamento discutidas em estudos como Belle et al. (2017), Onal e Kumkale (2022), Odetunde (2022) e Kim e Campbell (2025). Contudo, essa formação precisa estar inserida em sistemas mais estruturados, transparentes e participativos, como também sugerem estudos sobre desenho, execução, justiça e eficácia da avaliação, entre eles Kondrasuk (2011), Iqbal et al. (2015), McClendon et al. (2020), Micacchi et al. (2024) e Tarigan et al. (2025).

Dessa forma, a avaliação de desempenho deve ser compreendida como um processo organizacional contínuo, e não como um ato administrativo isolado. Quando os critérios são ambíguos, o feedback é insuficiente, as evidências não são registradas e os resultados não são utilizados de forma coerente, ampliam-se as condições para subjetividade, percepções de injustiça e distorções avaliativas. Logo, os resultados obtidos reforçam a necessidade de sistemas de avaliação mais claros, rastreáveis, participativos e orientados ao desenvolvimento.

5. CONCLUSÕES, CONTRIBUTOS, LIMITAÇÕES E INVESTIGAÇÃO FUTURA

A revisão sistemática desenvolvida permitiu reunir e analisar a literatura recente sobre erros, vieses, falhas estruturais e recomendações nos sistemas de avaliação de desempenho. A partir dos 37 artigos incluídos, verificou-se que a qualidade da avaliação resulta da articulação entre julgamento humano, desenho do sistema, critérios, evidências, feedback, calibração, cultura organizacional e finalidade institucional. Essa leitura afasta

uma explicação centrada apenas no avaliador e reforça a necessidade de compreender a avaliação como um processo organizacional integrado.

5.1. *Conclusões principais*

A revisão sistemática realizada permitiu responder à questão central deste trabalho, ao demonstrar que os erros, vieses e falhas estruturais mais recorrentes nos sistemas de avaliação de desempenho, entre 2010 e 2025, resultam da interação entre julgamento humano, desenho do sistema e condições organizacionais. A literatura analisada mostra que o julgamento humano continua vulnerável a distorções, mas também evidencia que essas distorções são favorecidas, ampliadas ou mantidas por sistemas mal desenhados, critérios pouco claros, feedback insuficiente, baixa documentação, reduzida calibração e fraca transparência na utilização dos resultados.

Quanto aos erros e vieses cognitivos mais frequentemente identificados, os resultados apontam para a recorrência da subjetividade do avaliador, da leniência, da severidade, do efeito *halo*, do efeito *horn*, da recência, da tendência central, do favoritismo, da política organizacional e outros vieses cognitivos ou contextuais associados à memória, à comparação, à confirmação, à ancoragem e às expectativas prévias. Estes problemas não aparecem com a mesma intensidade em todos os estudos, nem com o mesmo tipo de evidência. Alguns artigos oferecem evidência empírica direta, como Belle et al. (2017), Merkel et al. (2021), Onal e Kumkale (2022), Odetunde (2022) e Kim e Campbell (2025), enquanto outros discutem esses fenômenos de forma conceptual ou a partir de percepções dos participantes. Ainda assim, a recorrência dessas categorias confirma que os julgamentos avaliativos continuam vulneráveis a distorções que podem comprometer a precisão, a equidade e a utilidade da avaliação.

As falhas processuais, contextuais e estruturais aparecem como elemento decisivo para compreender por que esses erros persistem. A avaliação tende a perder consistência quando os critérios são vagos, quando os avaliadores não são preparados, quando as evidências de desempenho não são registadas ao longo do ciclo, quando o feedback é episódico ou pouco útil e quando os resultados são utilizados de forma pouco transparente. Acresce ainda que a avaliação não ocorre em ambiente neutro. Relações hierárquicas, normas culturais, redes informais, interesses políticos, expectativas prévias e percepções de justiça influenciam a forma como o desempenho é observado, interpretado e classificado. Assim, a fragilidade da avaliação não reside somente no julgamento em si, mas nas condições que permitem, orientam, documentam, calibram e tornam compreensível esse julgamento.

No que respeita às recomendações propostas pela literatura, os resultados mostram que as soluções mais recorrentes envolvem clarificação de critérios, metas e rubricas, fortalecimento do feedback contínuo e qualificado, treino e calibração dos avaliadores, maior transparência, participação do avaliado, documentação de evidências, uso criterioso de múltiplas fontes e apoio tecnológico para registo e rastreabilidade. Contudo, nenhuma dessas medidas deve ser interpretada como solução isolada. A formação dos avaliadores, por exemplo, é necessária, mas perde alcance quando os critérios permanecem ambíguos ou quando a organização não dispõe de mecanismos de revisão. Do mesmo modo, feedback contínuo, avaliação 360 graus e tecnologia só produzem ganhos quando integrados num sistema coerente, com finalidade clara, critérios partilhados e uso transparente dos resultados.

A distribuição das recomendações por níveis de análise reforça essa leitura integrada. No nível individual, a literatura valoriza a preparação do avaliador, a

consciência sobre vieses, a capacidade de observar comportamentos e a utilização consistente de evidências. No nível interpessoal, ganham relevo o feedback, a participação do avaliado, a confiança e a qualidade da comunicação. No nível organizacional, concentram-se medidas relacionadas com critérios claros, rubricas, documentação, calibração, transparência, justiça percebida e mecanismos de revisão. No nível estratégico, a avaliação surge associada ao alinhamento com objetivos organizacionais, ao desenvolvimento dos trabalhadores, à cultura de desempenho, à tecnologia de suporte e às decisões de gestão de recursos humanos. Dessa forma, a mitigação dos erros e vieses exige atuação integrada sobre pessoas, relações, processos, instrumentos e finalidade institucional.

A literatura recente também revela contributos, lacunas e tendências relevantes. O principal contributo está no deslocamento de uma leitura centrada apenas no erro individual para uma compreensão mais ampla da avaliação de desempenho como prática organizacional, social e cognitiva. A investigação recente tem dado maior atenção à justiça percebida, ao feedback contínuo, à cultura organizacional, à política interna, à digitalização e ao desenho dos sistemas avaliativos. Apesar disso, persistem lacunas importantes. Ainda há fragmentação conceptual, diferentes formas de nomear problemas semelhantes, diversidade metodológica elevada e pouca acumulação de evidência sobre a eficácia combinada dos mecanismos de mitigação. A literatura identifica muitas recomendações, mas ainda testa pouco, em contextos organizacionais reais, quais combinações produzem melhores resultados em termos de precisão avaliativa, justiça percebida, motivação para melhorar e desenvolvimento profissional.

5.2. *Contributos teóricos e implicações práticas*

Do ponto de vista teórico, este trabalho contribui ao organizar a literatura recente sobre erros e recomendações na avaliação de desempenho a partir de uma leitura integrada. A matriz de extração e análise permitiu distinguir erros, vieses, falhas estruturais e mecanismos de mitigação, evitando tratar todos esses elementos como se fossem fenómenos equivalentes. Essa distinção é relevante porque um viés pode explicar uma distorção no julgamento, enquanto uma falha estrutural pode criar as condições para que diferentes vieses se manifestem com maior intensidade. Logo, a contribuição não está apenas em listar problemas recorrentes, mas em mostrar como eles se articulam.

Outro contributo teórico decorre da articulação entre problemas e respostas possíveis. Ao relacionar falhas de desenho, subjetividade, feedback insuficiente, critérios pouco claros, contexto cultural, justiça percebida e mecanismos de mitigação, a revisão mostra que a avaliação de desempenho deve ser analisada como sistema. Essa perspectiva ajuda a superar abordagens fragmentadas, ao mostrar que o julgamento avaliativo é moldado por critérios, procedimentos, evidências, relações e finalidades institucionais.

Do ponto de vista prático, os resultados indicam que organizações interessadas em melhorar seus sistemas avaliativos devem começar pela arquitetura do processo. Critérios ambíguos, ausência de registos, feedback frágil, baixa calibração e pouca transparência tornam a avaliação mais dependente da memória, da percepção e das relações informais. Nessas condições, mesmo avaliadores bem-intencionados podem produzir julgamentos inconsistentes. Por isso, a melhoria da avaliação exige critérios observáveis, documentação ao longo do ciclo, formação dos avaliadores, reuniões de calibração, participação do avaliado, feedback orientado para a melhoria e mecanismos de revisão capazes de reforçar a confiança no sistema.

Também se destaca a necessidade de alinhar a avaliação com decisões de desenvolvimento e gestão de pessoas. Quando a avaliação é usada apenas como procedimento administrativo ou classificatório, perde parte de sua utilidade organizacional. Quando, por outro lado, é articulada com desenvolvimento, reconhecimento, planeamento de carreira, formação e melhoria do desempenho, tende a ganhar legitimidade. A avaliação de desempenho torna-se mais robusta quando deixa de ser um evento isolado e passa a funcionar como processo contínuo de acompanhamento, aprendizagem e responsabilização.

5.3. Limitações

Algumas limitações devem ser consideradas na interpretação dos resultados. A revisão foi delimitada ao período entre 2010 e 2025, aos idiomas inglês e português e às bases EBSCO, *Scopus* e *Web of Science*. Embora esses critérios tenham garantido foco, rastreabilidade e qualidade académica, podem ter deixado de fora estudos publicados noutras línguas, noutras bases ou em literatura cinzenta.

Outra limitação decorre da impossibilidade de recuperar integralmente dois artigos potencialmente relevantes. Como esses estudos não foram analisados na íntegra, não foi possível avaliar sua contribuição para o corpus final. Além disso, os 37 artigos incluídos apresentam naturezas metodológicas distintas, abrangendo estudos empíricos, conceptuais, documentais, revisões e abordagens mistas. Essa heterogeneidade enriquece a análise, mas impede tratar todas as evidências com o mesmo peso.

Por fim, a codificação das categorias envolveu interpretação analítica. Embora tenha sido baseada na leitura integral dos artigos e orientada pela matriz de extração e análise, a classificação de problemas, vieses, falhas e recomendações não é um

procedimento puramente mecânico. Trata-se de uma síntese qualitativa, sujeita aos limites próprios de qualquer leitura temática da literatura.

5.4. *Investigação futura*

Investigações futuras poderão aprofundar a eficácia dos mecanismos de mitigação em contextos organizacionais reais. Em especial, seriam úteis estudos longitudinais e experiências de campo capazes de analisar se a combinação entre formação de avaliadores, calibração, feedback contínuo, documentação de evidências e participação do avaliado reduz efetivamente distorções avaliativas e melhora as percepções de justiça.

Também se mostra relevante comparar diferentes setores, culturas e modelos organizacionais. A literatura analisada sugere que fatores culturais e institucionais influenciam a forma como os sistemas avaliativos são implementados, mas ainda há espaço para compreender melhor como essas diferenças afetam a aceitação, a legitimidade e a eficácia da avaliação.

Outra linha de investigação promissora diz respeito ao papel da tecnologia. Ferramentas digitais, sistemas de feedback em tempo real e soluções apoiadas por inteligência artificial podem favorecer registro, rastreabilidade e acompanhamento, mas também podem introduzir novos riscos de enviesamento, opacidade e dependência excessiva de métricas. Assim, a investigação futura deve examinar não apenas se a tecnologia torna a avaliação mais eficiente, mas em que condições ela contribui para avaliações mais justas, transparentes e orientadas ao desenvolvimento.

5.5. *Considerações finais*

A principal conclusão desta revisão é que a mitigação dos erros e vieses na avaliação de desempenho não depende apenas de corrigir o avaliador, mas de aperfeiçoar o sistema no qual o avaliador atua. Avaliações mais justas, precisas e úteis exigem coerência entre critérios, pessoas, processos, evidências, feedback, cultura e finalidade institucional. Quando esses elementos não estão alinhados, a avaliação tende a reproduzir subjetividade, inconsistência e desconfiança. Quando são estruturados de forma integrada, a avaliação pode cumprir melhor sua função de apoiar o desenvolvimento dos trabalhadores, fortalecer decisões de gestão e contribuir para organizações mais transparentes, responsáveis e orientadas à melhoria contínua.

REFERÊNCIAS BIBLIOGRÁFICAS

- Agbola, R. M., Hemans, S., & Abena, S. (2011). The effectiveness of performance appraisal as a tool for enhancing employee performance in the public health sector. *Global Management Journal*, 3(1/2), 81–96.
- Aguinis, H. (2014). *Performance management* (3rd ed.). Pearson.
- Alves, M. W. F. M., & Mariano, E. B. (2018). Climate justice and human development: A systematic literature review. *Journal of Cleaner Production*, 202, 360–375.
<https://doi.org/10.1016/j.jclepro.2018.08.091>
- Appelbaum, S. H., Roy, M., & Gilliland, T. (2011). Globalization of performance appraisals: Theory and applications. *Management Decision*, 49(4), 570–585.
<https://doi.org/10.1108/00251741111126495>
- Armstrong, M. (2015). *Armstrong's handbook of performance management: An evidence-based guide to delivering high performance* (5th ed.). Kogan Page.
- Aro-Gordon, S. (2016). Leveraging information technology for effective performance appraisal in the Nigerian public service. *SDMIMD Journal of Management*, 7(2), 21–39. <https://doi.org/10.15533/sdm/2016/v7i2/104322>
- Ayers, R. S. (2013). Building goal alignment in federal agencies' performance appraisal programs. *Public Personnel Management*, 42(4), 495–520.
<https://doi.org/10.1177/0091026013496077>
- Belle, N., Cantarelli, P., & Belardinelli, P. (2017). Cognitive biases in performance appraisal: Experimental evidence on anchoring and halo effects with public sector managers and employees. *Review of Public Personnel Administration*, 37(3), 275–294. <https://doi.org/10.1177/0734371X17704891>

- Bizzi, L. (2018). The problem of employees' network centrality and supervisors' error in performance appraisal: A multilevel theory. *Human Resource Management*, 57(2), 515–528. <https://doi.org/10.1002/hrm.21880>
- Botelho, C. (2024). The influence of performance-driven cultures on performance appraisal best practices effectiveness. *International Journal of Productivity and Performance Management*, 73(4), 1187–1211. <https://doi.org/10.1108/IJPPM-06-2022-0297>
- Briner, R., & Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool. In D. M. Rousseau (Ed.), *The Oxford handbook of evidence-based management* (pp. 112–129). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199763986.013.0007>
- Cammeraat, E., Lalowski, M., Ratcheva, V., & Zafar, A. (2021). *Drivers of productivity growth: An international perspective*. OECD Publishing.
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402. <https://doi.org/10.1002/asi.21525>
- Dangelico, R. M. (2016). Green product innovation: Where we are and where we are going. *Business Strategy and the Environment*, 25(8), 560–576. <https://doi.org/10.1002/bse.1886>
- Daud, Z., Elsaman, H. A., & Sallehuddin, M. R. (2023). The office politics error as a new dimension in performance appraisal implementations: A case study and conceptual model in Malaysian financial sector. *Acta Innovations*, 48, 48–60. <https://doi.org/10.32933/ActaInnovations.48.4>

- DeNisi, A. S., & Murphy, K. R. (2017). Performance appraisal and performance management: 100 years of progress? *Journal of Applied Psychology*, 102(3), 421–433. <https://doi.org/10.1037/apl0000085>
- Dohare, S., & Singh, N. K. (2021). A case study of performance appraisal system in public sector and private sector banks at Uttarakhand. *International Journal of Research in Commerce & Management*, 12(2), 5–8.
- Gough, D., Oliver, S., & Thomas, J. (Eds.). (2012). *An introduction to systematic reviews*. SAGE.
- Haque, M. A. (2012). Performance appraisal system of Bangladesh Civil Service: An analysis of its efficacy. *International Public Management Review*, 13(1), 38–59.
- Honarmand Shahzileh, Z., & Aghajan, A. M. (2015). Performance appraisal: A review and case study. *International Journal of Business Excellence*, 8(5), 584–600. <https://doi.org/10.1504/IJBEX.2015.071279>
- Iqbal, M. Z., Akbar, S., & Budhwar, P. (2015). Effectiveness of performance appraisal: An integrated framework. *International Journal of Management Reviews*, 17(4), 510–533. <https://doi.org/10.1111/ijmr.12050>
- Javidmehr, M., & Ebrahimpour, M. (2015). Performance appraisal bias and errors: The influences and consequences. *International Journal of Organizational Leadership*, 4(3), 286–302. https://ijol.cikd.ca/article_60464_e664300a56883a0e6b8a49a754f0bd81.pdf
- Kim, P., & Campbell, J. W. (2025). Halos, performance matching, and PSM: Experimental evidence of performance appraisal bias among Korean government workers. *Review of Public Personnel Administration*. Advance online publication. <https://doi.org/10.1177/0734371X251371241>

- Kondrasuk, J. N. (2011). So what would an ideal performance appraisal look like? *Journal of Applied Business and Economics*, 12(1), 57–71.
- Kromrei, H. (2015). Enhancing the annual performance appraisal process: Reducing biases and engaging employees through self-assessment. *Performance Improvement Quarterly*, 28(2), 53–64. <https://doi.org/10.1002/piq.21192>
- Kumari, S. (2012). A study on performance appraisal errors of telecom managers in Navi Mumbai region. *SIES Journal of Management*, 8(2), 3–14.
- Lameque, I. do C., Velez, M. J. P., & Botelho, C. M. D. A. (2023). Purposes of performance appraisal: A systematic review and agenda for future research. *International Journal of Professional Business Review*, 8(7), e02274. <https://doi.org/10.26668/businessreview/2023.v8i7.2274>
- Larsen, D. A. (2019). Using the elaboration likelihood model to explain performance appraisal inaccuracies. *Journal of Management Policy and Practice*, 20(4), 64–77.
- Liu, X., & Dong, K. (2012). Development of the civil servants' performance appraisal system in China: Challenges and improvements. *Review of Public Personnel Administration*, 32(2), 149–168. <https://doi.org/10.1177/0734371X12438244>
- Lunenburg, F. C. (2012). Performance appraisal: Methods and rating errors. *International Journal of Management, Business, and Administration*, 15(1), 1–5.
- Makhubela, D. K., & Malatji, T. L. (2024). Performance appraisal as a tool to achieve higher efficiency and productivity in the South African public sector, with reference to Limpopo Department of Education. *Perspectives of Law and Public Administration*, 13(1), 55–67. <https://doi.org/10.62768/PLPA/2024/13/1/06>

- McClendon, J. A., Deckop, J. R., Han, S., & Petrucci, T. (2020). A study of system execution of performance appraisal. *International Journal of Selection and Assessment*, 28(3), 322–336. <https://doi.org/10.1111/ijsa.12289>
- Merkel, S., Chan, H. F., Schmidt, S. L., & Torgler, B. (2021). Optimism and positivity biases in performance appraisal ratings: Empirical evidence from professional soccer. *Applied Psychology: An International Review*, 70(3), 1100–1127. <https://doi.org/10.1111/apps.12266>
- Micacchi, L., Vidé, F., Giacomelli, G., & Barbieri, M. (2024). Performance appraisal justice and employees' work engagement in the public sector: Making the most of performance appraisal design. *Public Administration*, 102(3), 815–840. <https://doi.org/10.1111/padm.12952>
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18(1), 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Murphy, K. R., & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Sage Publications.
- Murphy, K. R., & DeNisi, A. (2023). New approaches to dealing with performance management: Getting rid of performance appraisals is not the answer. *IIM Ranchi Journal of Management Studies*, 2(2), 143–158. <https://doi.org/10.1108/IRJMS-09-2023-0074>
- Odetunde, O. J. (2022). Revisiting performance appraisal measurement: Construction and validation of Rater Bias Scale. *Organizational Psychology*, 12(4), 123–143. <https://doi.org/10.17323/2312-5942-2022-12-4-123-143>

- Onal, I. G., & Kumkale, G. T. (2022). Effectiveness of source-monitoring training in reducing halo error and negativity bias in a performance appraisal setting. *Applied Psychology: An International Review*, 71(4), 1635–1653.
<https://doi.org/10.1111/apps.12350>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71.
<https://doi.org/10.1136/bmj.n71>
- Park, S. (2017). Motivating raters through work design: Applying the job characteristics model to the performance appraisal context. *Cogent Psychology*, 4(1), Article 1287320. <https://doi.org/10.1080/23311908.2017.1287320>
- Prati, F., Menegatti, M., Moscatelli, S., Kana Kenfack, C. S., Pireddu, S., Crocetti, E., Mariani, M. G., & Rubini, M. (2019). Are mixed-gender committees less biased toward female and male candidates? An investigation of competence-, morality-, and sociability-related terms in performance appraisal. *Journal of Language and Social Psychology*, 38(5/6), 586–605.
<https://doi.org/10.1177/0261927X19844808>
- Rivera, M., Jiang, C., & Kumar, S. (2024). Seek and ye shall find: An empirical examination of the effects of seeking real-time feedback on employee performance evaluations. *Information Systems Research*, 35(2), 783–806.
<https://doi.org/10.1287/isre.2021.0130>

- Rolle, C., & Klingner, D. (2012). Performance appraisal systems as a strategic human resource management tool in the Bahamian public service. *International Journal of Public Administration*, 35(5), 303–314.
<https://doi.org/10.1080/01900692.2012.655459>
- Rosales Sánchez, C., Díaz-Cabrera, D., & Hernández-Fernaund, E. (2019). Does effectiveness in performance appraisal improve with rater training? *PLOS ONE*, 14(9), Article e0222694. <https://doi.org/10.1371/journal.pone.0222694>
- Sanner, B., Evans, K., & Fernandez, D. (2022). Do desperate times call for desperate measures? The effect of crises on performance appraisals. *Human Performance*, 35(3/4), 218–240. <https://doi.org/10.1080/08959285.2022.2108034>
- Selvarajan, T. T., & Cloninger, P. A. (2012). Can performance appraisals motivate employees to improve performance? A Mexican study. *The International Journal of Human Resource Management*, 23(15), 3063–3084.
<https://doi.org/10.1080/09585192.2011.637069>
- Shim, D. C., Park, S., & Park, H. H. (2025). Linking performance appraisal and government employees' organizational citizenship behavior. *Review of Public Personnel Administration*. Advance online publication.
<https://doi.org/10.1177/0734371X241237564>
- Singh, V. K., Kochar, B., & Yüksel, S. (2010). An empirical study on the efficiency of performance appraisal system in Oil and Natural Gas Commission (ONGC), India. *İşletme Araştırmaları Dergisi*, 2(2), 65–78.
- Tarigan, A., Gustomo, A., & Bangun, Y. R. (2025). Are performance appraisals in the public sector fair? Exploring bias and best practices. *Cogent Business &*

Management, 12(1), Article 2458760.

<https://doi.org/10.1080/23311975.2025.2458760>

Tshukudu, T. T. (2020). Overcoming errors and biases in the Botswana public sector performance appraisal system: The case of Ministry of Trade and Industry.

Revue Européenne du Droit Social, 49(4), 101–108.

United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. United Nations. <https://sdgs.un.org/2030agenda>

Wilkinson, A., Redman, T., & Dundon, T. (2017). *Contemporary human resource management: Text and cases* (5th ed.). Pearson.

Yahiaoui, D., Nakhle, S. F., & Farndale, E. (2021). Culture and performance appraisal in multinational enterprises: Implementing French headquarters' practices in Middle East and North Africa subsidiaries. *Human Resource Management*, 60(5), 771–785. <https://doi.org/10.1002/hrm.22063>