



Universidade de Brasília

Efetividade das Avaliações de Desempenho nas Organizações

Dissertação de Mestrado apresentada ao Programa de Pós Graduação em Psicologia Social, do Trabalho e das Organizações como requisito à obtenção do grau de Mestre em Psicologia Social, do Trabalho e das Organizações.

Tatiana Andrade da Silva Lima

Orientação: Profa. Dra. Elaine Rabelo Neiva

Sumário

Apresentação Geral	07
Avaliação de Desempenho: Uma Revisão Sistemática da Produção Científica Global (2013-2025)	09
Resumo	10
Abstract	11
Introdução	12
Método	15
Resultados	21
Discussão	29
Conclusão	32
Contribuições do Estudo e Corroboração Teórica	32
Limitações do Estudo	33
Agenda de Pesquisa	33
Referências	35
Determinantes da Efetividade na Avaliação de Desempenho: Uma Síntese da Literatura sobre Antecedentes e Resultados Organizacionais	42
Resumo	43
Abstract	44
Introdução	45

	3
Método	47
Resultados	51
Discussão	56
Conclusão	60
Contribuições Teóricas	60
Implicações Práticas	61
Limitações do Estudo	61
Agenda de Pesquisa	61
Referências	63
Mensurando a Efetividade da Avaliação de Desempenho: Desenvolvimento de uma Escala para o Contexto Organizacional	69
Resumo	70
Abstract	71
Introdução	72
Método	74
Procedimentos Teóricos e Qualitativos	75
Construção do Instrumento e Validação de Conteúdo	76
Análise Fatorial Exploratória - AFE	77
Amostra	77
Procedimentos de coleta	77
Instrumento	78

Análise de Dados	78
Análise Fatorial Confirmatória - AFC	78
Amostra	78
Procedimentos de coleta	79
Instrumento	79
Análise de Dados	79
Validade e confiabilidade dos construtos	80
Resultados	81
Discussão	98
Conclusão	101
Contribuições do Estudo	101
Corroboração com a Literatura	101
Limitações do Estudo	102
Agenda de Pesquisa	102
Referências	103
Conclusão	108

Lista de Tabelas e Figuras

Avaliação de Desempenho: Uma Revisão Sistemática da Produção Científica Global
(2013-2025)

Tabela 1: Definição dos critérios de elegibilidade baseada no acrônimo PICOS 16

Tabela 2: Resultados brutos da estratégia de busca por base de dados 17

Figura 1: Fluxograma PRISMA de seleção dos estudos	18
Figura 2: Distribuição das Publicações por Base de Dados Bibliográfica	22
Figura 3: Evolução Temporal das Publicações	23
Figura 4: Distribuição Geográfica da Produção Científica (2013-2025)	24
Figura 5: Distribuição da Produção por Setor Organizacional	25
Figura 6: Distribuição dos Estudos por Abordagem Metodológica de Pesquisa	26
Determinantes da Efetividade na Avaliação de Desempenho: Uma Síntese da Literatura sobre Antecedentes e Resultados Organizacionais	
Tabela 1: Definição dos critérios de elegibilidade baseada no acrônimo PICOS	47
Tabela 2: Resultados brutos da estratégia de busca por base de dados	48
Figura 1: Fluxograma PRISMA de seleção dos estudos	49
Figura 2: Modelo Teórico-Conceitual dos Mecanismos de Efetividade da AD	50
Tabela 3: Frequência dos Determinantes Críticos de Efetividade	52
Tabela 4: Síntese das Estratégias de Treinamento de Avaliadores e Impacto na Efetividade	53
Figura 3: Modelo de Impacto da Efetividade da AD	56
Mensurando a Efetividade da Avaliação de Desempenho: Desenvolvimento de uma Escala para o Contexto Organizacional	
Tabela 1: Categorias analíticas de efetividade da AD emergentes dos grupos focais ...	81
Tabela 2: Cálculo do CVC dos Itens da Escala	83
Tabela 3: Caracterização da amostra	86

Tabela 4: Variância explicada com base nos autovalores e análise paralela	87
Tabela 5: Estatísticas de Ajuste para o Modelo Unifatorial	88
Tabela 6: Cargas Fatoriais da Matriz de Fator Único	89
Tabela 7: Caracterização da amostra	91
Tabela 8: Índices de ajuste do modelo testado	92
Tabela 9: Cargas fatoriais padronizadas	93
Figura 1: Modelo estrutural/path diagram da AFC	94
Figura 2: Gráfico de dispersão (Scatter plot) entre os escores do instrumento atual e a dimensão de crenças, apresentando a linha de tendência linear	95
Tabela 10: Itens da Escala de Efetividade da Avaliação de Desempenho (EEAD)	96

Apresentação Geral

A Avaliação de Desempenho (AD) constitui-se como uma das práticas mais estratégicas e, simultaneamente, mais complexas no âmbito da Gestão de Pessoas. Historicamente compreendida como uma ferramenta técnica de controle e mensuração de resultados, a AD tem passado por uma transição paradigmática nas últimas décadas. A literatura científica recente indica um deslocamento do foco estritamente métrico para uma perspectiva psicossocial, na qual a percepção de justiça, a qualidade das interações e o papel da liderança emergem como elementos determinantes para o sucesso do sistema.

Apesar de sua relevância, a implementação da AD nas organizações é frequentemente acompanhada por desafios significativos. Críticas relacionadas à burocratização do processo e à falta de utilidade percebida pelos trabalhadores sugerem que a eficiência técnica de um instrumento não garante, isoladamente, a sua efetividade organizacional. Nesse sentido, a efetividade da avaliação de desempenho passa a ser compreendida não apenas pelo alcance de metas, mas pela percepção de justiça, utilidade e suporte que o sistema oferece ao desenvolvimento do indivíduo e da organização.

Diante desse cenário, esta dissertação investiga os contornos teóricos e práticos da efetividade na avaliação de desempenho, estruturando-se em três etapas complementares que compõem o corpo deste trabalho.

O primeiro manuscrito apresenta uma revisão sistemática da produção científica global entre 2013 e 2025. O objetivo deste estudo foi mapear tendências metodológicas e temáticas, revelando uma hegemonia de abordagens quantitativas e a consolidação de

temas como justiça organizacional e liderança transformacional como preditores centrais do desempenho.

No segundo manuscrito aprofunda-se na investigação sobre os determinantes da efetividade da AD por meio de uma síntese integrativa da literatura. Este estudo identifica que a efetividade é um construto dependente da qualidade das trocas sociais, especialmente no que tange à justiça interacional e à precisão do *feedback*. Os resultados indicam que a efetividade percebida atua como uma variável mediadora essencial, convertendo insumos de gestão em resultados tangíveis, como o aumento do comprometimento e da produtividade.

Por fim, o terceiro manuscrito aborda a lacuna instrumental identificada nos estudos anteriores. Dada a escassez de ferramentas com indícios de validade para o contexto brasileiro que considerem a multidimensionalidade da efetividade, este estudo descreve o desenvolvimento e a busca de evidências de validade da Escala de Efetividade da Avaliação de Desempenho (EEAD). Por meio de um delineamento de métodos mistos e análises fatoriais (exploratória e confirmatória), o instrumento demonstrou robustez psicométrica, oferecendo uma contribuição prática e teórica para a mensuração do construto.

Em suma, esta dissertação busca não apenas compreender o estado atual da avaliação de desempenho, mas também fornecer subsídios científicos e instrumentais que auxiliem gestores e pesquisadores na promoção de sistemas de avaliação mais humanos, justos e efetivos.

**Avaliação de Desempenho: Uma Revisão Sistemática
da Produção Científica Global (2013-2025)**

*Performance Appraisal: A Systematic Review
of Global Scientific Production (2013-2025)*

Resumo

A avaliação de desempenho (AD) permanece como uma das práticas mais críticas e contestadas na gestão de recursos humanos. Este estudo apresenta uma revisão sistemática da literatura sobre a AD em contextos organizacionais, abrangendo o período de 2013 a 2025. O recorte temporal justifica-se pela análise da produção científica consolidada na última década, acrescida de publicações recentes para garantir a atualidade e a robustez das evidências. O objetivo foi identificar as tendências, metodologias, contextos e lacunas predominantes na produção científica da última década. Utilizando o protocolo PRISMA, foram analisados 275 artigos selecionados a partir de buscas no periódico CAPES. Os resultados revelam uma hegemonia de abordagens quantitativas (69,45%) e uma geografia de pesquisa deslocada para o continente asiático (42,18%), com destaque para países como Índia, Paquistão, China, Malásia e Indonésia, e foco nos setores de educação, saúde e indústria. Tematicamente, a literatura evidenciou a transição de uma visão puramente técnica para a valorização do processo interacional, destacando a percepção de justiça organizacional, a influência da política interna, o impacto da liderança transformacional e a emergência de sistemas eletrônicos (e-PA) e modelos matemáticos de decisão. Conclui-se que a aplicabilidade da AD não reside na sofisticação do instrumento, mas na qualidade das relações interpessoais, na percepção de justiça procedimental e na adaptação do sistema ao contexto cultural e organizacional.

Palavras-chave: Avaliação de Desempenho; Avaliação de Desempenho Eletrônica; Gestão de Pessoas; Gestão de Recursos Humanos; Revisão Sistemática.

Abstract

Performance appraisal (PA) remains one of the most critical and contested practices in human resource management. This study presents a systematic literature review on PA in organizational contexts, covering the period from 2013 to 2025. The chosen time frame is justified by the analysis of scientific production consolidated in the last decade, supplemented by recent publications to ensure the timeliness and robustness of the evidence. The objective was to identify the predominant trends, methodologies, contexts, and gaps in the scientific production of the last decade. Using the PRISMA protocol, 275 articles were analyzed, selected from searches in the CAPES journal. The results reveal a hegemony of quantitative approaches (69.45%) and a research geography shifted towards the Asian continent (42.18%), with emphasis on countries such as India, Pakistan, China, Malaysia, and Indonesia, and a focus on the education, health, and industry sectors. Thematically, the literature highlighted the transition from a purely technical view to the valorization of the interactional process, emphasizing the perception of organizational justice, the influence of internal politics, the impact of transformational leadership, and the emergence of electronic systems (e-PA) and mathematical decision models. It is concluded that the effectiveness of decision-making does not reside in the sophistication of the instrument, but in the quality of interpersonal relationships, the perception of procedural justice, and the adaptation of the system to the cultural and organizational context.

Keywords: Performance Appraisal; Electronic Performance Appraisal; People Management; Human Resources Management; Systematic Review.

Introdução

A Avaliação de Desempenho (AD) é definida como o processo sistemático de mensuração e análise do comportamento profissional e dos resultados alcançados pelos trabalhadores em relação aos objetivos organizacionais (Longo, 2007). Historicamente, esse processo constitui um dos subsistemas mais críticos e controversos da Gestão de Recursos Humanos (GRH). Para além de sua função original de controle, a AD consolidou-se como uma ferramenta estratégica voltada ao desenvolvimento humano, ao alinhamento cultural e ao fortalecimento do compromisso organizacional (Girdharwal, 2019). A análise da evolução desse campo revelou que, ao longo de duas décadas, os temas de pesquisa amadureceram de uma visão puramente técnica para uma abordagem que busca compreender as tendências globais e as lacunas teóricas persistentes (Francis et al., 2021). Esse amadurecimento reflete a transição da AD de um rito burocrático para um mecanismo que, em contextos de alta complexidade e instabilidade, pode assumir funções sociais e psicossociais inesperadas.

Diferente da lógica tradicional de fiscalização, pesquisas em cenários de crise demonstram que a AD pode ser ressignificada como uma ferramenta de suporte e proteção da equipe. Em contextos de conflito armado, por exemplo, as lógicas ocidentais de controle e motivação cedem espaço para uma “lógica de paz organizacional”, em que o sistema de avaliação funciona como um refúgio para criar harmonia interna e garantir a sobrevivência emocional do grupo, muitas vezes resultando em uma inflação sistemática de notas para evitar estresse adicional em profissionais já traumatizados (Giangreco et al., 2012). Essa plasticidade do sistema demonstra que a AD não é um instrumento estático, mas um processo que se adapta às necessidades sociopolíticas da organização e de seus membros.

A transformação digital também impôs novas dinâmicas à AD por meio da implementação de sistemas eletrônicos (Curzi et al., 2019), conhecidos como e-PA (*Electronic Performance Appraisal*). Essa transição para o ambiente digital, embora prometa maior transparência e agilidade, exige uma análise cuidadosa de como as plataformas tecnológicas interagem com as variáveis processuais da gestão. Em setores críticos como o hospitalar, a adoção do e-PA é vista como um caminho para a sustentabilidade social organizacional, mas sua aplicação prática revela que a estrutura tecnológica deve estar profundamente alinhada ao planejamento das sessões e à qualidade do diálogo entre avaliador e avaliado para que não se torne apenas uma automação da subjetividade (Ullah, Álvarez-Otero et al., 2021). Assim, o foco contemporâneo da pesquisa desloca-se da simples existência de um sistema para a compreensão de como o processo (incluindo o planejamento e os múltiplos insumos de informação) molda a experiência do trabalhador e sua relação com a organização (Girdharwal, 2019).

No entanto, a literatura científica contemporânea aponta para uma persistente lacuna entre a intenção estratégica da AD e a sua prática cotidiana (Schleicher et al., 2018; Brown et al., 2019). Essa discrepância, observada recorrentemente em estudos teóricos e empíricos, motiva a necessidade de investigar como a produção científica tem evoluído e quais lacunas ainda persistem. Esta prática ainda é frequentemente marcada por percepções de injustiça e burocracia excessiva (Bodart et al., 2016; Neiva, 2020), além de uma acentuada vulnerabilidade a enviesamentos políticos (DeNisi & Murphy, 2017; Ibrahim et al., 2016). A introdução de metodologias mais abrangentes, como o sistema de *feedback* 360 graus, buscou mitigar esses problemas ao incluir múltiplas fontes de avaliação (superiores, subordinados e pares). Contudo, estudos em ambientes acadêmicos revelam que esse método ainda enfrenta desafios significativos,

especialmente devido às grandes discrepâncias entre a autoavaliação dos gestores e a percepção de seus superiores, evidenciando que a subjetividade humana permanece como um elemento central e desafiador no processo avaliativo (Görün et al., 2018).

A literatura sobre o tema revela que ainda existe uma fragmentação nas abordagens de pesquisa: por um lado, uma vertente focada em aspectos psicométricos busca reduzir o erro de avaliação e por outro, emerge uma perspectiva socioconstrutivista que privilegia a interação discursiva, em que o processo comunicativo é frequentemente mais determinante para os resultados finais do que o sistema tecnológico propriamente dito (Peixoto & Caetano, 2013; Schleicher et al., 2018). Adicionalmente, observa-se uma tensão geopolítica e cultural na produção do conhecimento. Enquanto modelos teóricos dominantes foram desenvolvidos no contexto, principalmente, da Europa Ocidental e América do Norte, a última década testemunhou um aumento de pesquisas empíricas oriundas de economias emergentes da Ásia e da África, que trazem à tona idiosincrasias culturais, como a distância do poder e o coletivismo, capazes de moderar ou até inverter as relações esperadas entre as práticas de avaliação e o desempenho individual (Kagama & Irungu, 2018).

Nesse cenário, este estudo justifica-se pela necessidade de mapear a extensão e a natureza dessa produção científica recente, problematizando as contradições inerentes à prática da avaliação de desempenho. A ausência de uma síntese que integre essas variáveis humanas, técnicas e contextuais fundamenta a realização desta revisão sistemática, que busca consolidar o estado atual sobre a AD e fornecer subsídios para o avanço da teoria organizacional no horizonte de 2013 a 2025. O problema central reside na constatação de que, apesar da sofisticação crescente das ferramentas, incluindo inteligência artificial e algoritmos matemáticos avançados (Ojokoh et al., 2019), as organizações continuam a enfrentar desafios relacionados à precisão das avaliações e ao

efetivo engajamento. Estudos recentes indicam que a AD é um construto social complexo, influenciado por variáveis contextuais de país e setor (Kagama & Irungu, 2018; Ibrahim et al., 2016), variáveis processuais, como a qualidade do *feedback* e o treinamento (Schleicher et al., 2018; Girdharwal, 2019), e variáveis psicológicas, como a motivação e o contrato psicológico (Aguinis, 2013; Sanyal & Biswash, 2014).

Diante do exposto, esta revisão sistemática tem como objetivo mapear e analisar a produção científica sobre a Avaliação de Desempenho (AD) no período de 2013 a 2025, identificando como as variáveis processuais, psicológicas e contextuais interagem na configuração de sistemas de gestão de desempenho efetivos. Para guiar esta investigação, formulou-se a seguinte pergunta de pesquisa: “Quais as principais tendências, lacunas e evidências apresentadas pela literatura científica recente sobre os fatores que influenciam a aplicabilidade da avaliação de desempenho em diferentes contextos organizacionais?”.

Método

A presente pesquisa caracteriza-se como uma revisão sistemática da literatura, conduzida conforme as diretrizes do protocolo PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), revisado por Page et al. (2022). Este delineamento foi selecionado por permitir a identificação, seleção, avaliação e síntese de evidências científicas de forma objetiva, replicável e transparente, minimizando vieses (Galvão & Pereira, 2014). Para garantir o rigor metodológico e a reprodutibilidade do estudo, a estratégia de pesquisa foi estruturada com base no acrônimo PICOS, conforme detalhado na Tabela 1.

Com a questão norteadora: “Quais são as características, as abordagens metodológicas e os principais achados reportados na literatura científica sobre sistemas

de avaliação de desempenho organizacional entre os anos de 2013 e 2025?”, foram incluídos artigos científicos originais, publicados em periódicos e com revisão por pares. A escolha do período justifica-se pela necessidade de capturar a evolução da AD na última década. Quanto ao idioma, foram selecionados estudos em inglês, português e espanhol, permitindo uma análise abrangente que integra o panorama global. O escopo temático limitou-se a estudos que explorassem a AD como um construto social e sistema de gestão, priorizando análises sobre comportamento organizacional, justiça e processos comunicativos.

Tabela 1

Definição dos critérios de elegibilidade baseada no acrônimo PICOS.

Acrônimo	Componente	Critérios de Inclusão
P	População	Trabalhadores, gestores e profissionais inseridos em contextos organizacionais (públicos ou privados) de diversos setores (saúde, educação, indústria, etc.).
I	Intervenção	Sistemas, processos, ferramentas ou lógicas de Avaliação de Desempenho (AD), incluindo métodos tradicionais, <i>feedback</i> 360 graus e sistemas eletrônicos (e-PA).
C	Comparação	Não foi exigido um grupo de comparação específico, mas foram considerados estudos que comparam diferentes fontes de avaliação (ex: autoavaliação vs. superior) ou sistemas manuais vs. digitais.
O	Desfechos (<i>Outcomes</i>)	Características dos sistemas, determinantes de aplicabilidade, percepção de justiça, impactos no comprometimento organizacional, satisfação no trabalho e barreiras de implementação.
S	Delineamento (<i>Study Design</i>)	Estudos empíricos originais (quantitativos, qualitativos ou de métodos mistos) publicados em periódicos.

Nota. e-PA = electronic Performance Appraisal.

A busca foi realizada no portal de periódicos CAPES, o qual viabilizou o acesso integrado às bases de dados *Web of Science*, EBSCO e PubMed, bem como ao acervo de relevantes editoras acadêmicas, com destaque para Elsevier, Emerald Publishing, SAGE, Springer Nature, Taylor & Francis e Wiley. A estratégia utilizou os descritores

“*PERFORMANCE APPRAISAL*” (título) e “*ORGANIZATION*” (qualquer campo), combinados pelo operador booleano *AND* (Galvão & Pereira, 2014).

Tabela 2

Resultados brutos da estratégia de busca por base de dados.

Base de Dados	Filtros e Estratégia de Busca	Resultados Brutos	Duplicatas Removidas	Total para Triagem
Portal CAPES	<i>PERFORMANCE APPRAISAL</i> (título) <i>AND ORGANIZATION</i> (qualquer campo); Filtros: 2013-2025, Artigo, Acesso Aberto, Revisado por Pares.	198	-	198
Web of Science	<i>PERFORMANCE APPRAISAL</i> (título) <i>AND ORGANIZATION</i> (qualquer campo); Filtro: 2013-2025.	260	49	211
EBSCO	<i>PERFORMANCE APPRAISAL</i> (título) <i>AND ORGANIZATION</i> (qualquer campo); Filtro: 2013-2025.	55	23	32
PubMed	<i>PERFORMANCE APPRAISAL</i> (título) <i>AND ORGANIZATION</i> (qualquer campo); Filtro: 2013-2025.	26	11	15
TOTAL		539	83	456

Nota. Embora o levantamento bibliográfico tenha abrangido um total de 50 bases de dados e editoras científicas acessadas via portal de periódicos CAPES, a tabela apresenta apenas as fontes que se destacaram pelo maior volume de artigos recuperados.

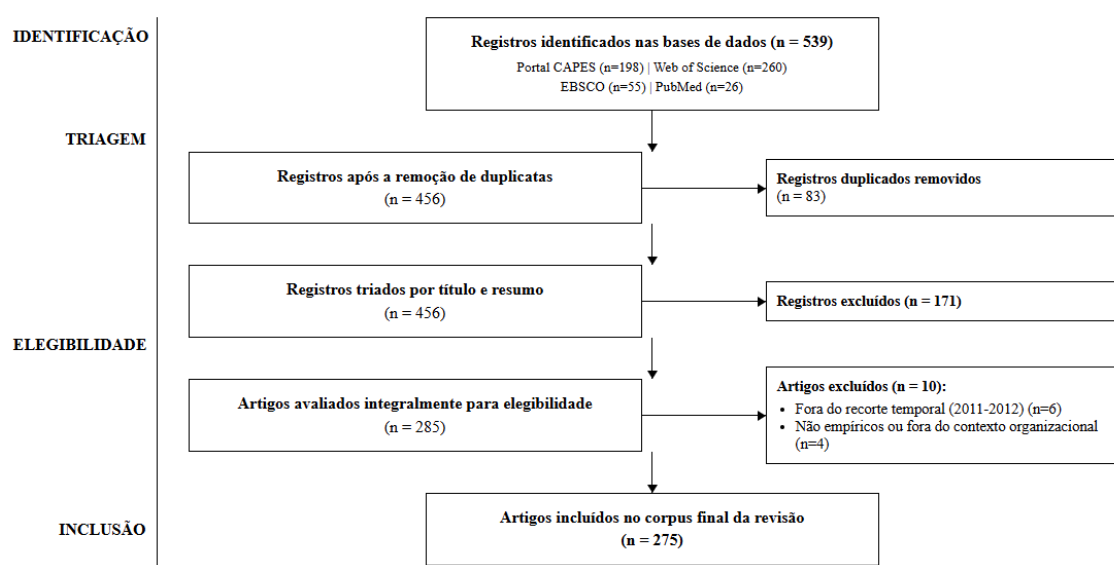
O levantamento inicial resultou em um total bruto de 539 registros. A remoção de duplicatas foi realizada no software Microsoft Excel, utilizando a funcionalidade de formatação condicional para identificação de valores repetidos, resultando na exclusão de 83 artigos (49 na *Web of Science*, 23 na EBSCO e 11 na PubMed). Após essa etapa, 456 registros foram submetidos à triagem inicial.

O processo de seleção ocorreu em duas fases: na primeira, a leitura de títulos e resumos levou à exclusão de 171 registros que não atendiam aos critérios de inclusão (fora do tema ou não organizacionais), restando 285 artigos para análise detalhada. Na

segunda fase, procedeu-se à leitura integral dos textos. Nesta etapa, 10 artigos foram excluídos: 6 por pertencerem aos anos de 2011 e 2012 (fora do recorte temporal de 2013-2025) e 4 por não serem estudos empíricos ou focarem exclusivamente em desempenho financeiro. Ao final, a amostra definitiva foi composta por 275 artigos.

Figura 1

Fluxograma PRISMA de seleção dos estudos.



Nota. Adaptado de Page et al. (2022).

A extração de dados da amostra final de 275 artigos foi conduzida sistematicamente por meio de uma matriz de categorização padronizada no Microsoft Excel, em conformidade com o preconizado pelo item 10 do protocolo PRISMA (Page et al., 2022). Esta matriz foi estruturada para garantir a captura organizada de variáveis essenciais, divididas em quatro eixos principais: (1) identificação e contexto, abrangendo o título, autoria, ano de publicação, periódico e país de origem; (2) delineamento metodológico, registrando a abordagem (quantitativa, qualitativa ou de métodos mistos), a natureza (transversal ou longitudinal) e os procedimentos específicos

de coleta, tais como surveys, estudos de caso, análises bibliométricas e relatos de experiência; (3) escopo da amostra, detalhando a descrição dos participantes, os instrumentos de medida utilizados e a natureza das organizações estudadas, contemplando os setores público, privado e o terceiro setor; e (4) conteúdo técnico, consolidando os objetivos, as variáveis analisadas (independentes, dependentes e mediadoras), os principais resultados encontrados, as limitações e as lacunas sugeridas para pesquisas futuras. Esse processo de extração assegurou a consistência das evidências que fundamentam a integração de estudos de naturezas diversas, desde a aplicação de sistemas multi-rater em universidades (Görün et al., 2018) até a implementação de plataformas digitais em hospitais (Ullah, Ahmad et al., 2021).

A análise e a integração dos dados extraídos seguiram a abordagem de síntese narrativa, conforme as diretrizes do protocolo PRISMA (Page et al., 2022) para revisões em que a meta-análise estatística não se mostra aplicável (Galvão & Pereira, 2014). A síntese narrativa foi adotada devido ao fato de: (1) os estudos incluídos apresentarem naturezas distintas, variando de análises bibliométricas globais (Francis et al., 2021), modelos matemáticos (Esen et al., 2016), estudos de caso em zonas de conflito armado (Giangreco et al., 2012) e levantamentos estatísticos em diversos setores, o que impede a utilização de uma unidade métrica comum para uma agregação puramente estatística; e (2) a abordagem permitir integrar evidências de naturezas diversas, unindo achados qualitativos, como percepções e discursos, a dados quantitativos de escores de desempenho e correlações (Girdharwal, 2019). O processamento dos resultados ocorreu por meio do agrupamento dos achados em categorias temáticas emergentes da literatura, organizadas em eixos de discussão.

Para a etapa de extração e sistematização das informações, utilizou-se uma planilha eletrônica padronizada para organizar os dados dos artigos selecionados em

dimensões complementares. Inicialmente, coletaram-se os metadados de identificação, como o número de registro, a base de dados de origem, o título do estudo, o DOI e o ano de publicação. Em seguida, procedeu-se ao detalhamento do contexto e do delineamento metodológico, registrando-se o país de realização da pesquisa, o delineamento adotado, a abordagem e a natureza do estudo. Essa dimensão incluiu ainda a caracterização dos participantes, a descrição dos instrumentos de medida empregados e a classificação das variáveis em independentes, dependentes ou indiretas.

Na sequência, a extração focou no conteúdo e nos achados das obras, resumizando os objetivos declarados, os principais resultados alcançados e as contribuições específicas para o campo da Avaliação de Desempenho. Por fim, realizou-se uma análise voltada às lacunas e limitações reportadas pelos próprios autores, integrando sugestões para pesquisas futuras, considerações sobre a qualidade do manuscrito e a justificativa final para sua inclusão na revisão. Essa estrutura de coleta fundamentou a avaliação do risco de viés, permitindo identificar, a partir dos dados extraídos, aspectos como o tamanho e a representatividade das amostras, a utilização de delineamentos exclusivamente transversais, que limitam inferências causais, e a dependência de medidas de autorrelato passíveis de induzir ao viés de variância comum do método.

A presente revisão integra o projeto de pesquisa intitulado “Efetividade das Avaliações de Desempenho nas Organizações”, vinculado ao Instituto de Psicologia da Universidade de Brasília (IP/UnB). O projeto macro foi submetido e aprovado pelo Comitê de Ética em Pesquisa (CEP), estando registrado sob o Parecer nº 8.050.720. A condução deste estudo seguiu rigorosamente as diretrizes éticas vigentes para pesquisas científicas, assegurando a transparência e o compromisso com as normas de integridade acadêmica.

Resultados

Nesta seção, apresentam-se os achados derivados da revisão sistemática, organizados em duas dimensões principais. Inicialmente, detalha-se o perfil bibliométrico e metodológico da amostra, descrevendo a distribuição temporal das publicações, as fontes de informação, a origem geográfica dos estudos e as abordagens predominantes. Em um segundo momento, procede-se à síntese categórica dos conteúdos, na qual as evidências são discutidas a partir de eixos temáticos emergentes: o impacto da transformação digital e dos sistemas eletrônicos (*e-PA*), as percepções de justiça e variáveis psicossociais no processo avaliativo, e a influência das variáveis contextuais e culturais na gestão de desempenho.

A execução da estratégia de busca e o processo de seleção, fundamentados no protocolo PRISMA (Page et al., 2022), resultaram em um *corpus* final de 275 estudos. Esta amostra oferece um panorama detalhado da produção científica sobre Avaliação de Desempenho (AD) no período compreendido entre 2013 e 2025. Esta delimitação se justifica pela necessidade de capturar a produção científica consolidada na última década, marcada pela transição definitiva da AD para modelos baseados em transformação digital e sistemas eletrônicos (*e-PA*). Enquanto revisões clássicas e de grande impacto, como as de DeNisi e Murphy (2017) e Schleicher et al. (2018), sistematizaram o progresso histórico e teórico do campo até meados da década de 2010, o presente estudo busca preencher a lacuna subsequente.

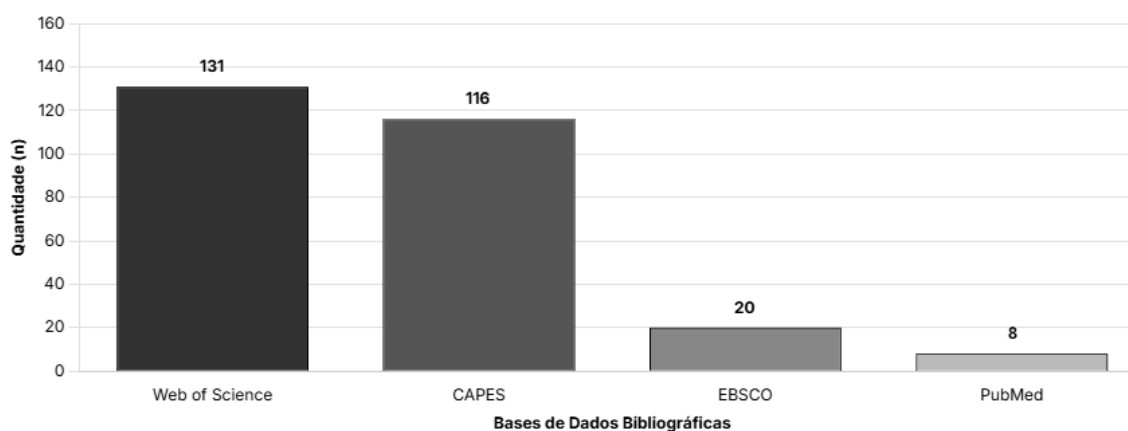
Diferente de outras revisões que focaram na transição do paradigma psicométrico para o de gestão de desempenho, o intervalo 2013-2025 permite analisar a maturação de novas variáveis intervenientes, como a sustentabilidade social organizacional e o impacto de algoritmos e inteligência artificial nos processos

avaliativos. Assim, ao comparar este levantamento com revisões anteriores como as de DeNisi e Murphy (2017) e Schleicher et al. (2018), observou-se que este estudo atua na integração de evidências recentes que refletem as mudanças drásticas nos contextos de trabalho contemporâneos, incluindo o trabalho remoto e as lógicas de paz organizacional em cenários de crise, que não foram plenamente abarcadas por sínteses teóricas da década passada.

A análise descritiva dos artigos selecionados revela uma transformação significativa na geografia e no perfil da produção científica sobre AD. No que tange às fontes de informação, os registros foram recuperados por meio de diferentes bases de dados e plataformas de busca acadêmica. A *Web of Science* apresentou a maior concentração, com 47,64% da amostra, seguida pelas buscas realizadas via portal de periódicos CAPES (42,18%), que atuou como a principal interface de acesso a múltiplos repositórios e coleções licenciadas. De forma complementar, as plataformas EBSCO e PubMed contribuíram com participações de 7,27% e 2,91%, respectivamente.

Figura 2

Distribuição das Publicações por Base de Dados Bibliográfica.

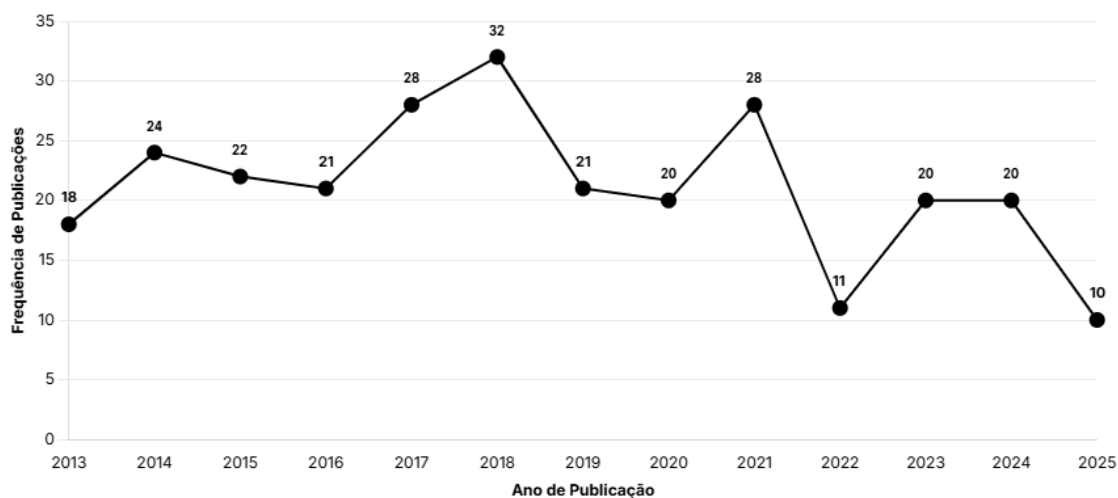


Nota. N = 275. A base *Web of Science* (n = 131) e o portal CAPES (n = 116) representam as fontes primárias de evidência científica para o tema no período analisado.

Cronologicamente, a produção científica apresenta-se distribuída de forma consistente ao longo da última década, com picos de publicação identificados nos anos de 2017 a 2018 e 2021. Notou-se, ainda, a manutenção do interesse acadêmico nos anos mais recentes, com 7,27% das publicações registradas em cada um dos anos de 2023 e 2024.

Figura 3

Evolução Temporal das Publicações.



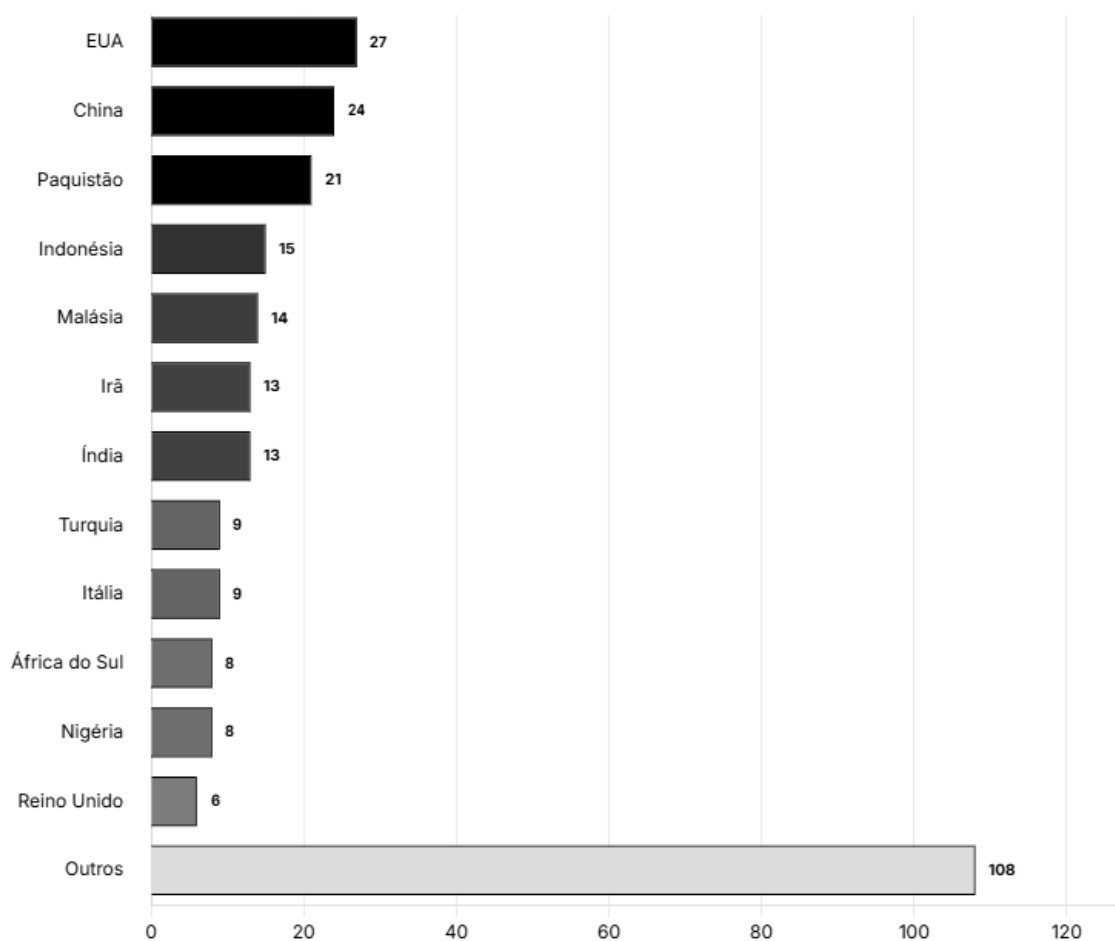
Nota. Frequência anual de artigos identificados nas bases *Web of Science*, CAPES, EBSCO e PubMed. Os valores acima dos pontos indicam a contagem absoluta de publicações por ano. Observa-se o ápice de produção em 2018 (n = 32).

Quanto à distribuição geográfica, embora os modelos teóricos tenham raízes em contextos Europeus e Norte Americanos, a última década evidencia um deslocamento

do protagonismo para as economias emergentes. Individualmente, os Estados Unidos lideram a produção com 9,82% dos estudos ($n = 27$), seguidos pela China com 8,73% ($n = 24$). Contudo, o bloco de países emergentes apresenta uma densidade expressiva, com destaque para o Paquistão ($n = 21$), Indonésia ($n = 15$), Malásia ($n = 14$), Índia ($n = 13$) e Irã ($n = 13$). Observou-se, simultaneamente, um volume relevante de pesquisas oriundas da Turquia ($n = 9$), Itália ($n = 9$), África do Sul ($n = 8$), Nigéria ($n = 8$) e Reino Unido ($n = 6$). Os demais países e estudos com múltiplas nacionalidades, frequências menores ou não discriminadas, agrupados na categoria “outros”, representam 39,27% da amostra ($n=108$).

Figura 4

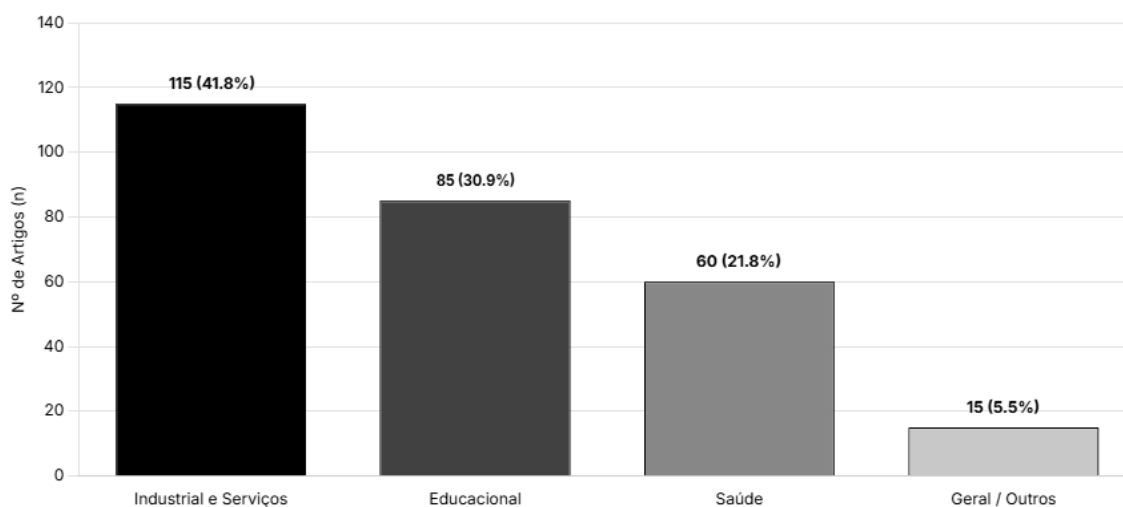
Distribuição Geográfica da Produção Científica (2013-2025).



Setorialmente, a amostra concentra-se em três grandes frentes que demonstram a versatilidade e a criticidade da AD em diferentes cenários. O Setor Educacional destaca-se pela gestão docente e cargos de chefia acadêmica, analisando a AD como indutor de qualidade pedagógica (Görün et al., 2018; Kagema & Irungu, 2018). O Setor de Saúde apresenta a avaliação como um pilar de sustentabilidade social e segurança do paciente, buscando reduzir a opacidade em ambientes hospitalares (Pereira & Moreira, 2015; Ullah, Álvarez-Otero et al., 2021). Por fim, o Setor Industrial e de Serviços foca na eficiência produtiva e na retenção de talentos em mercados competitivos, como o bancário e o automotivo (Alqahtani, 2024; Sanyal & Biswash, 2014).

Figura 5

Distribuição da Produção por Setor Organizacional.



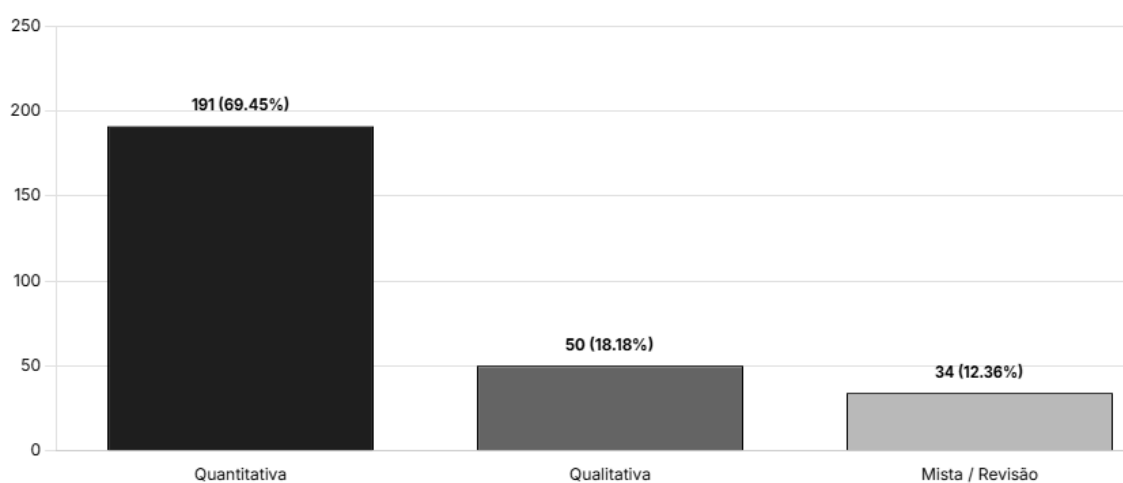
Nota. A amostra demonstra a versatilidade da AD, com predominância no Setor Industrial e de Serviços ($n = 115$) e expressiva relevância nos campos Educacional ($n = 85$) e de Saúde ($n = 60$).

No âmbito metodológico, observa-se uma hegemonia da abordagem Quantitativa, presente em 69,45% da amostra, caracterizada majoritariamente pelo uso

de *surveys* e modelagens estatísticas. A abordagem Qualitativa representa 18,18% dos estudos, sendo fundamental para explorar a subjetividade das interações humanas durante as sessões de *feedback*. Nota-se, contudo, uma tendência crescente de pesquisas que utilizam modelagens matemáticas sofisticadas, como a Teoria dos Conjuntos Aproximados (*Rough Sets Theory*) (Ojokoh et al., 2019) e a Lógica Fuzzy (Arogundade et al., 2020; Esen et al., 2016; Khalilzadeh et al., 2020; Kurniawan et al., 2024; Samuel et al., 2014; Shafii et al., 2016; Suranti & Ikhsan, 2017), visando mitigar o viés humano no processo decisório, identificando e compreendendo como os vieses afetam a precisão da AD.

Figura 6

Distribuição dos Estudos por Abordagem Metodológica de Pesquisa.



Quanto à natureza temporal, verificou-se que 69,82% dos estudos possuem caráter transversal, capturando percepções em um único ponto no tempo. Essa predominância ressalta uma lacuna de estudos longitudinais, que permitiriam acompanhar a evolução dos efeitos da AD ao longo de ciclos plurianuais. Os

delineamentos de pesquisa mais frequentes incluem o Correlacional (6,55%) e o Estudo de Caso (2,91%).

A análise dos artigos selecionados permitiu a identificação de eixos temáticos que sintetizam a AD. A definição dos eixos temáticos foi realizada por meio de uma análise de conteúdo categórica, baseada na técnica de síntese temática. Este processo se iniciou durante a etapa de extração de dados, na qual as colunas “objetivos”, “principais resultados” e “considerações sobre o artigo” da planilha de sistematização foram comparadas. A partir dessa leitura, identificaram-se padrões recorrentes e convergências de conteúdo entre os estudos, que foram inicialmente agrupados por similaridade semântica.

Posteriormente, esses agrupamentos foram refinados e consolidados em eixos centrais, que refletem a complexidade do campo ao integrar variáveis técnicas, humanas e contextuais. A lógica de criação desses eixos considerou a distinção teórica entre a estrutura formal da avaliação (o “sistema”) e a qualidade de sua execução prática (o “processo”), permitindo organizar as evidências de forma que as interdependências entre a tecnologia, o comportamento humano e as questões culturais fossem evidenciadas.

A análise da literatura revelou uma distinção central entre a estrutura formal da avaliação (o “sistema”) e a qualidade de sua execução prática (o “processo”).

Observou-se que as facetas ligadas ao processo, como a utilidade percebida pelo trabalhador, a acurácia das observações e a preparação das sessões, estão mais associadas à satisfação do avaliado do que a sofisticação tecnológica dos formulários (Rodrigues et al., 2023). Quando esses elementos processuais são negligenciados, o sistema é frequentemente percebido como um rito burocrático politizado (configura-se

como um processo de avaliação reduzido ao cumprimento de formalidades administrativas desprovidas de propósito estratégico, cujos resultados são distorcidos por jogos de poder, favoritismos e interesses particulares em detrimento da objetividade técnica.) o que pode elevar a intenção de *turnover*, a menos que haja um forte ajuste entre o indivíduo e os valores da organização (Alqahtani, 2024).

Nesse cenário, a aceitação da AD está intrinsecamente ligada à percepção de justiça organizacional (Mok & Leong, 2021). A justiça interacional, que é a forma como o gestor trata o subordinado durante a conversa, é frequentemente mais valorizada do que a própria classificação final, especialmente em contextos acadêmicos (Kivipõld et al., 2020). O papel da liderança, portanto, atua como um mediador crítico. Já a liderança transformacional, em particular, exerce uma influência positiva ao converter o momento da avaliação em uma oportunidade de desenvolvimento e suporte social (Ullah, Álvarez-Otero et al., 2021). Sob essa ótica, o *feedback* deixa de ser uma entrega pontual e passa a ser explorado como um processo contínuo de comunicação (Ock & Oswald, 2015). A motivação para trabalhar é impulsionada quando o avaliador demonstra competência no suporte contínuo, o que reduz comportamentos defensivos e a aversão à crítica (Marchegiani et al., 2016; Sanyal & Biswash, 2014).

Quanto às configurações estruturais, a literatura destaca a implementação de sistemas eletrônicos de Avaliação de Desempenho, conhecidos como e-PA (*Electronic Performance Appraisal*), como uma via fundamental para a promoção da transparência organizacional. Estes sistemas compreendem plataformas digitais e softwares integrados, como os módulos de gestão de talentos em sistemas ERP (*Enterprise Resource Planning*) ou soluções em nuvem, que automatizam o fluxo de recolha, processamento e armazenamento de dados de desempenho. A justificativa para a sua adoção reside na necessidade de garantir a auditabilidade dos registros e minimizar

critérios puramente subjetivos através de uma padronização rigorosa, permitindo que a tecnologia atue como um elemento neutro no processo (Ullah, Álvarez-Otero et al., 2021).

O efeito desses sistemas sobre os principais desfechos organizacionais se manifesta, sobretudo, na mitigação de vieses que comprometem a integridade da avaliação. Este controle é operacionalizado através da arquitetura do software, que estabelece campos obrigatórios, exige evidências documentais para justificar notas extremas e, em modelos mais avançados, utiliza lógica *fuzzy* ou inteligência artificial para monitorizar padrões anômalos de pontuação. Dessa forma, o sistema consegue identificar e alertar sobre distorções como o efeito *halo*, onde uma característica positiva contamina todas as outras notas, ou a tendência central, onde o avaliador evita notas extremas para fugir de conflitos.

Tais evidências foram sustentadas por estudos, como os de Ullah, Ahmad et al. (2021), que analisaram o setor da saúde em economias emergentes. Nestas organizações com recursos escassos (termo que define instituições com orçamentos limitados, baixa maturidade digital e infraestruturas de RH precárias), o e-PA exerce um papel de “árbitro neutro”. Em contextos onde o nepotismo e as agendas políticas internas são predominantes, a percepção de que os dados são processados por uma plataforma centralizada e automatizada eleva significativamente a confiança do trabalhador na justiça procedimental, reduzindo o sentimento de desamparo perante a vontade subjetiva do supervisor imediato e fortalecendo o compromisso com a organização.

A adoção de múltiplas fontes de avaliação (sendo as mais tradicionais a autoavaliação, avaliação pelo superior hierárquico, avaliação dos pares, avaliação 180 e 360 graus) surge como uma tentativa de mitigar o viés de uma fonte única. Entretanto,

as evidências revelam que essa configuração pode gerar tensões devido às discrepâncias recorrentes entre a autoavaliação do gestor e a percepção de seus superiores, o que demanda estratégias de monitoramento para reduzir distorções como os erros de *halo* e negatividade (Görün et al., 2018; Onal & Kumkale, 2022).

Por fim, as variáveis culturais moderam significativamente essas relações. Em países com altos índices de desconfiança institucional, a AD enfrenta elevados níveis de rejeição devido ao viés percebido nos supervisores (Kagama & Irungu, 2018). Em contrapartida, emerge o conceito de “paz organizacional” em contextos de crise extrema, nos quais o sistema de avaliação deixa de ser um instrumento técnico para atuar como um suporte psicossocial, priorizando a harmonia interna e o acolhimento da equipe em detrimento da cobrança por metas (Giangreco et al., 2012).

A síntese dos estudos permitiu identificar fragilidades metodológicas recorrentes: (1) a hegemonia de delineamentos transversais, que impedem a verificação da causalidade a longo prazo; (2) o viés de autorrelato (*Common Method Bias*), presente em pesquisas que utilizam apenas percepções subjetivas dos respondentes, o que pode inflar as correlações estatísticas; e (3) a especificidade dos estudos carecem de replicação em outros contextos.

Discussão

A presente revisão sistemática, ao analisar a literatura produzida entre 2013 e 2025, revela um campo em transformação, em que as certezas estritamente técnicas dão lugar a uma compreensão diversificada e socialmente situada do fenômeno da Avaliação de Desempenho (AD). Os achados sugerem que a AD é um fenômeno complexo, no qual a técnica, representada por formulários e escalas, é secundária aos processos sociais e psicológicos que ocorrem no cotidiano organizacional.

A evidência mais forte desta pesquisa é a supremacia do “processo” sobre o “sistema”. Enquanto a literatura clássica concentrava esforços no aperfeiçoamento de formulários, as evidências atuais demonstram que a sofisticação instrumental torna-se irrelevante sem uma boa comunicação. Observou-se que as facetas ligadas ao processo, como o planejamento do *feedback* e a qualidade da interação, são preditores mais robustos do compromisso afetivo do que o sistema tecnológico em si (Girdharwal, 2019). Nota-se uma insatisfação generalizada quando os funcionários percebem a AD como um exercício puramente burocrático, desconectado de recompensas ou desenvolvimento real. Assim, discute-se que o sucesso da ferramenta depende menos da precisão psicométrica e mais da “experiência do trabalhador”, exigindo que a AD seja um processo contínuo de alinhamento e não apenas um evento punitivo anual.

O surgimento de estudos sobre o *Electronic Performance Appraisal* (e-PA), particularmente em economias emergentes, aponta para a tecnologia não apenas como ferramenta de eficiência, mas de legitimação. Em contextos de baixa confiança institucional ou viés de supervisão, como o reportado no Quênia (Kagama & Irungu, 2018), a digitalização oferece uma objetividade percebida que restaura a justiça procedimental. O sistema mediado por algoritmos, lógica *fuzzy* ou inteligência artificial pode ser percebido como mais confiável do que o julgamento humano, que pode estar influenciado por interesses políticos ou vieses de leniência (Ibrahim et al., 2016). A tecnologia atua, portanto, como um “árbitro neutro”, especialmente em setores críticos como o hospitalar e o de alta precisão técnica (Becker et al., 2017; Ullah, Ahmad et al., 2021).

Um ponto crítico na discussão contemporânea reside na constatação de que a AD é um processo social e político. Verificou-se que gestores podem inflar notas para evitar conflitos ou favorecer aliados, o que compromete a integridade do sistema

(Ibrahim et al., 2016). Além disso, a literatura alerta para a persistência de vieses de gênero e raça, especialmente em critérios de avaliação subjetivos, que podem desfavorecer minorias e inibir comportamentos inovadores (Leyerzapf et al., 2015). Erros clássicos, como o efeito *halo* e a tendência central, continuam sendo desafios práticos que exigem treinamentos focados não apenas na técnica de preenchimento, mas na conscientização sobre processos cognitivos e tomadas de decisão multivariadas.

A relação entre a AD e o desempenho final do trabalhador não é direta, sendo mediada por variáveis como motivação, engajamento e percepção de justiça. Confirmou-se que a liderança transformacional exerce um papel vital ao converter a avaliação em um evento de desenvolvimento, aumentando a sustentabilidade social organizacional (Ullah, Álvarez-Otero et al., 2021). No entanto, o contexto cultural desafia essas práticas pois em culturas onde há uma grande distância entre chefes e subordinados o *feedback* 360 graus ou a participação direta do subordinado podem gerar desconforto e resistências (Görün et al., 2018). Em cenários extremos, como na Palestina, a AD assume uma função de “paz organizacional”, priorizando a harmonia da equipe e o suporte psicossocial em detrimento da cobrança por metas de eficiência técnica (Giangreco et al., 2012).

Os resultados desta revisão reforçam a necessidade de integrar a AD aos demais subsistemas de Gestão de Pessoas, confirmando que a ferramenta não deve operar de forma isolada. Ao contrastar estes achados com revisões da última década, como as de DeNisi e Murphy (2017) e Schleicher et al. (2018), percebe-se que a lacuna entre a intenção estratégica e a prática cotidiana permanece como um desafio central na área. Todavia, enquanto os estudos anteriores focavam predominantemente na transição do paradigma psicométrico para o de gestão, esta revisão identifica que a digitalização (e-PA) introduziu uma nova camada de complexidade: a tecnologia pode conferir

transparência, mas, sem o suporte da liderança transformacional, corre o risco de apenas automatizar a subjetividade, conforme sugerido pelas evidências de Ullah, Álvarez-Otero et al. (2021). Na prática, recomenda-se que as organizações priorizem a transparência e a capacitação de gestores para o diálogo, preterindo softwares meramente complexos em favor de uma “tecnologia social” de comunicação, valorizando o sentido atribuído ao processo pelas pessoas envolvidas.

Apesar do rigor metodológico do presente levantamento, algumas limitações inerentes à literatura analisada devem ser consideradas. A predominância de delineamentos transversais (59,27%) e a forte dependência de medidas de autorrelato nas pesquisas primárias restringem a compreensão da evolução das percepções ao longo do tempo e podem inflar as correlações estatísticas devido ao viés de variância comum do método. Além disso, muitos achados indicam que as evidências ainda carecem de maior volume de replicações para garantir uma validade externa universal.

Para pesquisas futuras, é necessário que se adotem delineamentos longitudinais capazes de verificar a causalidade e a estabilidade das relações entre as variáveis de AD a longo prazo. Recomenda-se ainda que a agenda científica avance na análise de vertentes menos exploradas, como o impacto de processos avaliativos mal estruturados no estresse ocupacional e no *burnout*. Por fim, sugere-se investigar o impacto de novas metodologias de gestão ágil, como os OKR (*Objectives and Key Results*), e o papel da inteligência artificial na mitigação da subjetividade humana, buscando compreender como estas inovações tecnológicas e metodológicas reconfiguram o contrato psicológico e o engajamento no ambiente de trabalho contemporâneo.

Conclusão

Contribuições do Estudo e Corroboração Teórica

Esta revisão demonstrou que a Avaliação de Desempenho (AD) atravessa um período de transição, deixando de ser um rito burocrático de controle para se consolidar como um processo dinâmico de desenvolvimento e interação social. A principal contribuição deste estudo reside no mapeamento da natureza multidimensional da AD, integrando as variáveis técnicas, humanas e contextuais que moldam sua eficácia. Os resultados corroboram a literatura clássica e contemporânea (Schleicher et al., 2018; Peixoto & Caetano, 2013) ao reafirmar que o sucesso da avaliação não reside na sofisticação técnica dos instrumentos, mas na percepção de justiça organizacional e na qualidade do diálogo entre avaliador e avaliado. No horizonte de 2025, a AD afirma-se como uma ponte essencial para o desenvolvimento sustentável das pessoas, desde que a “automação da subjetividade” via sistemas digitais seja mediada por práticas humanizadas.

Limitações do Estudo

Apesar do rigor metodológico, esta pesquisa apresenta limitações. O recorte temporal e a seleção de bases de dados podem ter omitido estudos de nicho ou literatura cinzenta (teses e dissertações não publicadas em periódicos) que tragam perspectivas alternativas. Além disso, embora se tenha observado um crescimento de pesquisas em economias emergentes, a produção acadêmica ainda é fortemente influenciada por modelos teóricos norte-americanos e europeus, o que pode limitar a generalização das conclusões para contextos culturais com lógicas de trabalho distintas.

Agenda de Pesquisa

Para o avanço da teoria organizacional, propõe-se uma agenda de pesquisa que se concentre, primeiramente, na investigação do impacto da inteligência artificial e da análise de dados automatizada, avaliando como o uso de algoritmos na tomada de

decisão influencia o contrato psicológico e a percepção de viés por parte dos trabalhadores. Adicionalmente, sugere-se que futuras investigações contemplem a relação entre os sistemas de avaliação e a saúde mental, analisando as consequências longitudinais da mensuração constante de desempenho em contextos de alta pressão e de trabalho remoto. Por fim, incentiva-se a realização de estudos transculturais que aprofundem a compreensão sobre como variáveis como a distância do poder e o coletivismo moderam a recepção do *feedback* e a eficácia da AD em diferentes regiões globais, permitindo uma visão mais inclusiva e menos eurocêntrica do fenômeno.

Em suma, a AD deve ser compreendida como um construto social complexo que exige um alinhamento contínuo entre a intenção estratégica da organização e a experiência vivida pelo trabalhador.

Referências

- Alqahtani, M. (2024). Person-organization fit and organizational commitment: Mediating role between performance appraisal effectiveness and employee turnover intention. *Business Ethics and Leadership*, 8(3), 55–70.
[https://doi.org/10.61093/bel.8\(3\).55-70.2024](https://doi.org/10.61093/bel.8(3).55-70.2024)
- Arogundade, O. T., Ojokoh, B., Asogbon, M. G., Samuel, O. W., & Adeniyi, B. S. (2020). Fuzzy driven decision support system for enhanced employee performance appraisal. *International Journal of Human Capital and Information Technology Professionals*, 11(1), 17–30.
<https://doi.org/10.4018/IJHCITP.2020010102>
- Becker, J., Bridge, P., Brown, E., Ferrari-Anderson, J., & Lusk, R. (2017). The development and implementation of a performance appraisal framework for radiation therapists in planning and simulation. *Journal of Medical Radiation Sciences*, 64(4), 321–327. <https://doi.org/10.1002/jmrs.249>
- Bodart, C., Ferreira, L., & Santos, G. (2016). Avaliação de desempenho: Uma análise crítica a partir da perspectiva dos avaliados. *Revista Foco*, 8(2), 103–115.
https://doi.org/10.28950/1981-223x_revistafocoadm/2015.v8i2.132
- Brown, T. C., O’Kane, P., Mazumdar, B., & McCracken, M. (2019). Performance management: A scoping review of the literature and an agenda for future research. *Human Resource Development Review*, 18(1), 47–82.
<https://doi.org/10.1177/1534484318798533>

- Curzi, Y., Fabbri, T., Scapolan, A. C., & Boscolo, S. (2019). Performance appraisal and innovative behavior in the digital era. *Frontiers in Psychology, 10*, Artigo 1659. <https://doi.org/10.3389/fpsyg.2019.01659>
- 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>
- Esen, H., Hatipoğlu, T., & Boyacı, A. İ. (2016). A fuzzy approach for performance appraisal: The evaluation of a purchasing specialist. In J. J. Merelo, A. Rosa, J. M. Cadenas, A. Dourado, K. Madani, & J. Filipe (Eds.), *Computational intelligence* (pp. 235–251). Springer. https://doi.org/10.1007/978-3-319-26393-9_14
- Francis, B. R., Ahmad, R. bin, & Abdullah, S. M. binti. (2021). A bibliometric analysis on performance appraisal. *International Journal of Academic Research in Business and Social Sciences, 11*(9), 1797–1818. <http://dx.doi.org/10.6007/IJARBS/v11-i9/11253>
- Galvão, T. F., & Pereira, M. G. (2014). Revisões sistemáticas da literatura: passos para sua elaboração. *Epidemiologia e Serviços de Saúde, 23*(1), 183–184. <https://doi.org/10.5123/S1679-49742014000100018>
- Giangreco, A., Carugati, A., Sebastiano, A., & Al Tamimi, H. (2012). War outside, ceasefire inside: An analysis of the performance appraisal system of a public hospital in a zone of conflict. *Evaluation and Program Planning, 35*, 161–170. <https://doi.org/10.1016/j.evalprogplan.2010.11.004>

- Girdharwal, N. (2019). A comparative study of performance appraisal on effectiveness and organizational commitment. *Prabandhan: Indian Journal of Management*, 12(11), 47–54. <https://doi.org/10.17010/pijom/2019/v12i11/148409>
- Görün, M., Kayar, İ., & Varol, B. (2018). 360-degree performance appraisal and feedback system: A study with heads of departments in Çanakkale Onsekiz Mart University. *Gaziantep University Journal of Social Sciences*, 17(4), 1425–1437. <https://doi.org/10.21547/jss.449154>
- Ibrahim, Z., Ismail, A., Mohamed, N. A. K., & Raduan, N. S. M. (2016). Association of managers' political interests towards employees' feelings of distributive justice and job satisfaction in performance appraisal system. *Procedia - Social and Behavioral Sciences*, 224, 523–530. <https://doi.org/10.1016/j.sbspro.2016.05.429>
- Kagama, J., & Irungu, C. (2018). An analysis of teacher performance appraisals and their influence on teacher performance in secondary schools in Kenya. *International Journal of Education*, 11(1), 93–98. <https://doi.org/10.17509/ije.v11i1.11148>
- Kivipõld, K., Türk, K., & Kivipõld, L. (2020). Performance appraisal, justice and organizational effectiveness: A comparison between two universities. *International Journal of Productivity and Performance Management*, 70(1), 87–108. <https://doi.org/10.1108/IJPPM-05-2019-0229>
- Khalilzadeh, M., Abouhamzeh, G., Sapauskas, J., & Banaitis, A. (2020). Analysis of risk factors of performance appraisal methods with fuzzy weighted axiomatic design. *Transformations in Business & Economics*, 19(3), 159–174.

- Kurniawan, D., Al-Faqih, H., & Raisy, L. W. (2024). Development of a comprehensive performance appraisal instrument using Behaviorally Anchored Rating Scales and Fuzzy TOPSIS. *E3S Web of Conferences*, 484, Artigo 01007.
<https://doi.org/10.1051/e3sconf/202448401007>
- Leyerzapf, H., Abma, T. A., Steenwijk, R. R., Croiset, G., & Verdonk, P. (2015). Standing out and moving up: Performance appraisal of cultural minority physicians. *Advances in Health Sciences Education*, 20(4), 995–1010.
<https://doi.org/10.1007/s10459-014-9577-6>
- Longo, F. (2007). *Mérito e flexibilidade: A gestão das pessoas no setor público* (A. Corbisier, L. Jahn, L. R. Gil, & P. A. Barbosa, Trans.). Edições Fundap. (Obra original publicada em 2004)
- Marchegiani, L., Reggiani, T., & Rizzolli, M. (2016). Loss averse agents and lenient supervisors in performance appraisal. *Journal of Economic Behavior & Organization*, 131, 183–197. <https://doi.org/10.1016/j.jebo.2016.07.022>
- Mok, M. K.-M., & Leong, Y.-Y. (2021). Factors affecting the effectiveness of employees' performance appraisal in private hospitals in Malaysia. *International Journal of Business and Society*, 22(1), 257–275.
<https://doi.org/10.33736/ijbs.3174.2021>
- Neiva, E. R. (2020). Gestão e avaliação de desempenho no setor público brasileiro: Aspectos críticos e discussão sobre as práticas correntes. *Cadernos da Reforma Administrativa, FONACATE*, 6, 1–25.
- Ojokoh, B., Akinsulire, V., & Isinkaye, F. (2019). An automated implementation of academic staff performance evaluation system based on rough sets theory.

Australasian Journal of Information Systems, 23.

<https://doi.org/10.3127/ajis.v23i0.2033>

Ock, J., & Oswald, F. L. (2015). Managing the interpersonal aspect of performance management. *Industrial and Organizational Psychology*, 8(1), 111–119.

<https://doi.org/10.1017/iop.2015.8>

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, 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. (2022). A declaração PRISMA 2020: diretriz atualizada para relatar revisões sistemáticas. *Revista*

Panamericana de Salud Pública, 46, e112.

<https://doi.org/10.26633/RPSP.2022.112>

Peixoto, A. L. A., & Caetano, A. (2013). Avaliação de desempenho. In L. Borges & M.

Mourão (Orgs.). *O trabalho e as organizações* (pp. 528-556). Porto Alegre:

Artmed.

Pereira, N. M. D., & Moreira, V. (2015). Avaliação de desempenho dos profissionais de

enfermagem: Percepção de justiça dos avaliados. *Pensar Enfermagem*, 19(2),

2–18.

Rodrigues, R. I., Gomes, C., & Junça-Silva, A. (2023). The role of satisfaction with the

performance appraisal: A comparative study between the public and private

sectors. *Revista Brasileira de Gestão de Negócios*, 25(1), 72–87.

<https://doi.org/10.7819/rbgn.v25i1.4216>

Samuel, O. W., Omisore, M. O., & Atajeromavwo, E. J. (2014). Online fuzzy based decision support system for human resource performance appraisal.

Measurement, 55, 452–461. <https://doi.org/10.1016/j.measurement.2014.05.024>

Sanyal, M. K., & Biswash, S. B. (2014). Employee motivation from performance appraisal implications: Test of a theory in the software industry in West Bengal (India). *Procedia Economics and Finance*, 11, 182–196.

[https://doi.org/10.1016/S2212-5671\(14\)00187-7](https://doi.org/10.1016/S2212-5671(14)00187-7)

Schleicher, D. J., Baumann, H. M., Sullivan, D. W., Levy, P. E., Hargrove, D. C., & Barros-Rivera, B. A. (2018). Putting the system into performance management systems: A review and agenda for performance management research. *Journal of Management*, 44(6), 2209–2245. <https://doi.org/10.1177/0149206318755303>

Shafii, M., Hosseini, S. M., Arab, M., Asgharizadeh, E., & Farzianpour, F. (2016). Performance analysis of hospital managers using fuzzy AHP and fuzzy TOPSIS: Iranian experience. *Global Journal of Health Science*, 8(2), 137–147.

<http://doi.org/10.5539/gjhs.v8n2p137>

Suranti, D., & Ikhsan, M. (2017). Performance appraisal of Seluma districts agricultural extensionist with fuzzy simple additive weighting method. *Scientific Journal of Informatics*, 4(2), 169–178. <https://doi.org/10.15294/sji.v4i2.12016>

Ullah, Z., Ahmad, N., Scholz, M., Ahmed, B., Ahmad, I., & Usman, M. (2021).

Perceived accuracy of electronic performance appraisal systems: The case of a

non-for-profit organization from an emerging economy. *Sustainability*, 13(4),
Artigo 2109. <https://doi.org/10.3390/su13042109>

Ullah, Z., Álvarez-Otero, S., Sulaiman, M. A. B. A., Sial, M. S., Ahmad, N., Scholz,
M., & Omhand, K. (2021). Achieving organizational social sustainability
through electronic performance appraisal systems: The moderating influence of
transformational leadership. *Sustainability*, 13(10), Artigo 5611.
<https://doi.org/10.3390/su13105611>

**Determinantes da Efetividade na Avaliação de Desempenho:
Uma Síntese da Literatura sobre Antecedentes e Resultados Organizacionais**

*Determinants of Effectiveness in Performance Appraisal:
A Synthesis of the Literature on Organizational Antecedents and Outcomes*

Resumo

Esta revisão sistemática de literatura, de natureza integrativa, analisa os determinantes críticos e os resultados organizacionais associados à efetividade dos sistemas de Avaliação de Desempenho (AD). Fundamentada no protocolo PRISMA, a pesquisa sintetiza evidências de 73 artigos, identificando que a efetividade da AD transcende a precisão técnica dos instrumentos para se consolidar como um construto psicossocial dependente da qualidade das interações humanas. Os achados revelam uma cadeia causal onde facetas do processo, especialmente a justiça interacional e a qualidade do feedback, emergem como preditores mais robustos da aceitação do sistema do que o seu *design* estrutural. O estudo destaca ainda o papel vital da liderança e de treinamentos específicos, como o monitoramento de fontes, na mitigação de vieses cognitivos e na redução de distorções propositais. Conclui-se que a efetividade percebida atua como uma variável mediadora central, convertendo insumos técnicos e suporte gerencial em resultados tangíveis, como o aumento do comprometimento organizacional, da motivação e da produtividade. O trabalho oferece uma síntese funcional que orienta gestores de Recursos Humanos a priorizarem a transparência processual e o suporte contínuo à liderança, transformando a avaliação em um pilar estratégico para o bem-estar e o desempenho institucional.

Palavras-chave: Avaliação de Desempenho; Efetividade Organizacional; Justiça Organizacional; *Feedback*; Revisão Sistemática.

Abstract

This integrative systematic literature review analyzes the critical determinants and organizational outcomes associated with the effectiveness of Performance Appraisal (PA) systems. Based on the PRISMA protocol, the research synthesizes evidence from 73 articles, identifying that the effectiveness of PA transcends the technical precision of the instruments to consolidate itself as a psychosocial construct dependent on the quality of human interactions. The findings reveal a causal chain where facets of the process, especially interactional justice and feedback quality, emerge as more robust predictors of system acceptance than its structural design. The study also highlights the vital role of leadership and specific training, such as source monitoring, in mitigating cognitive biases and reducing intentional distortions. It concludes that perceived effectiveness acts as a central mediating variable, converting technical inputs and managerial support into tangible results, such as increased organizational commitment, motivation, and productivity. This work offers a functional synthesis that guides Human Resources managers to prioritize procedural transparency and continuous support for leadership, transforming evaluation into a strategic pillar for well-being and institutional performance.

Keywords: Performance Appraisal; Organizational Effectiveness; Organizational Justice; Feedback; Systematic Review.

Introdução

A efetividade dos sistemas de Avaliação de Desempenho (AD) é um tema central na gestão estratégica de pessoas (Sahu et al., 2014). Conceitualmente, a AD é definida como um processo formal e sistemático de identificação, observação, mensuração e desenvolvimento do desempenho humano nas organizações, visando alinhar as competências individuais aos objetivos estratégicos da instituição (Cascio, 2019). Embora algumas organizações dediquem esforços ao desenho de ferramentas e métricas sofisticadas, a transição da intenção organizacional para a efetividade real é frequentemente mediada por percepções individuais e dinâmicas sociais (Roberts, 1992).

Enquanto as abordagens tradicionais focaram na precisão das escalas, a discussão científica contemporânea se voltou para a compreensão dos determinantes da efetividade, aprofundando como os aspectos do sistema e as interações do processo convergem para gerar valor organizacional (Martin & Bartol, 1998).

A efetividade na AD é compreendida como um construto multidimensional que engloba a utilidade percebida, a satisfação dos usuários e a capacidade do sistema em promover mudanças comportamentais tangíveis (Nazaruddin et al., 2022). Alguns estudos indicam que um determinante para o sucesso dessa prática reside na percepção de justiça. Quando o processo é percebido como justo em suas dimensões, se cria uma base sólida para o fortalecimento da aceitação do sistema pelos trabalhadores (Abdullah & Malik, 2022). Além disso, a clareza nas políticas de implementação e o alinhamento com a cultura organizacional são apontados como fatores que atribuem maior robustez aos sistemas de gestão de desempenho (Panda et al., 2018).

Com relação ao processo, a qualidade da interação humana durante o ciclo avaliativo, especialmente no *feedback*, surge como um antecedente fundamental para a

efetividade. As pesquisas clássicas destacam que a habilidade de transformar a avaliação em um diálogo focado no desenvolvimento é um dos principais impulsionadores de sucesso do programa (Meyer & Walker, 1961). Em contrapartida, influências implícitas, como distorções propositais ou atrasos na comunicação dos resultados, podem comprometer a efetividade, transformando sistemas tecnicamente bem desenhados em rituais burocráticos (Benedict & Levine, 1988).

No nível dos resultados organizacionais, a literatura aponta que a AD, quando é efetiva, atua como um incentivador para o comprometimento organizacional e para a melhoria do desempenho geral da força de trabalho (Girdharwal, 2019). A junção de propósitos estratégicos e de desenvolvimento permite que a avaliação funcione como um preditor da efetividade organizacional, auxiliando na retenção de talentos e na motivação contínua (Sahu et al., 2014).

Apesar da vasta produção científica acumulada ao longo das últimas décadas, observou-se ainda uma fragmentação na compreensão de como os determinantes de *design* (fatores do sistema) e de execução (fatores do processo) se articulam para garantir resultados sustentáveis. A influência de culturas orientadas para o desempenho sobre a efetividade das melhores práticas ainda carece de sínteses que integrem achados de diferentes contextos globais (Botelho, 2024).

Diante do exposto, este estudo busca preencher a lacuna sobre como os elementos de *design* e as dinâmicas de execução interagem para consolidar a efetividade da AD. Para nortear esta investigação, formulou-se a seguinte questão de pesquisa: “Quais são os determinantes críticos (antecedentes) e os desfechos organizacionais (consequentes) que caracterizam a efetividade dos sistemas de avaliação de desempenho na literatura científica?”. Para responder a essa interrogação, o presente estudo tem como objetivo analisar as evidências disponíveis sobre os mecanismos que sustentam o

sucesso da AD. Buscou-se investigar as relações entre as variáveis do sistema e do processo, a fim de propor uma síntese funcional que auxilie gestores e pesquisadores na compreensão dos elementos fundamentais para a promoção da efetividade organizacional.

Método

A pesquisa caracteriza-se como uma revisão sistemática da literatura de natureza integrativa, conduzida conforme as diretrizes do protocolo PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), revisado por Page et al. (2022). Este delineamento foi selecionado por permitir a identificação, seleção e síntese de evidências sobre os mecanismos que sustentam a efetividade da Avaliação de Desempenho (AD) de forma objetiva, replicável e transparente (Galvão & Pereira, 2014). Seguindo as recomendações de Günther (2004) e as diretrizes do protocolo PRISMA (Page et al., 2022), o processo foi estruturado em cinco etapas fundamentais: (1) formulação da questão norteadora por meio da estratégia PICOS e definição dos critérios de elegibilidade; (2) estratégia de busca e seleção em bases de dados; (3) triagem de títulos e resumos para refinamento de originalidade; (4) extração e categorização de determinantes e resultados; e (5) síntese narrativa dos mecanismos de sucesso identificados.

A questão norteadora foi estruturada pelo acrônimo PICOS (Page et al., 2022): “Quais são os determinantes críticos e os resultados organizacionais associados à efetividade dos sistemas de avaliação de desempenho organizacional reportados na literatura científica?”. Este estudo focou especificamente em artigos que se desdobraram nos fatores que levam ao sucesso do sistema e as consequências geradas por uma AD efetiva.

Tabela 1

Definição dos critérios de elegibilidade baseada no acrônimo PICOS.

Acrônimo	Componente	Crítérios de Inclusão
P	População	Organizações de diversos setores (público, privado e terceiro setor) e seus membros (gestores, trabalhadores e docentes).
I	Intervenção	Sistemas e processos de Avaliação de Desempenho (AD), incluindo <i>design</i> do sistema, critérios de avaliação e práticas de <i>feedback</i> .
C	Comparação	Não aplicável (N/A).
O	Desfechos (<i>Outcomes</i>)	Determinantes da efetividade (ex: justiça percebida, suporte da gestão, qualidade do <i>feedback</i>) e resultados organizacionais (ex: comprometimento, motivação e produtividade).
S	Delineamento (<i>Study Design</i>)	Estudos empíricos (quantitativos, qualitativos e mistos), revisões de literatura e estudos de caso com foco analítico em efetividade.

A busca foi realizada no portal de periódicos CAPES, o qual viabilizou o acesso integrado às bases de dados *Web of Science*, EBSCO e PubMed, bem como ao acervo de relevantes editoras acadêmicas, com destaque para Elsevier, Emerald Publishing, SAGE, Springer Nature, Taylor & Francis e Wiley.. A estratégia utilizou os descritores “*PERFORMANCE APPRAISAL*” (título) e “*EFFECTIVENESS*” (título), combinados pelo operador booleano *AND* (Galvão & Pereira, 2014).

Tabela 2

Resultados brutos da estratégia de busca por base de dados.

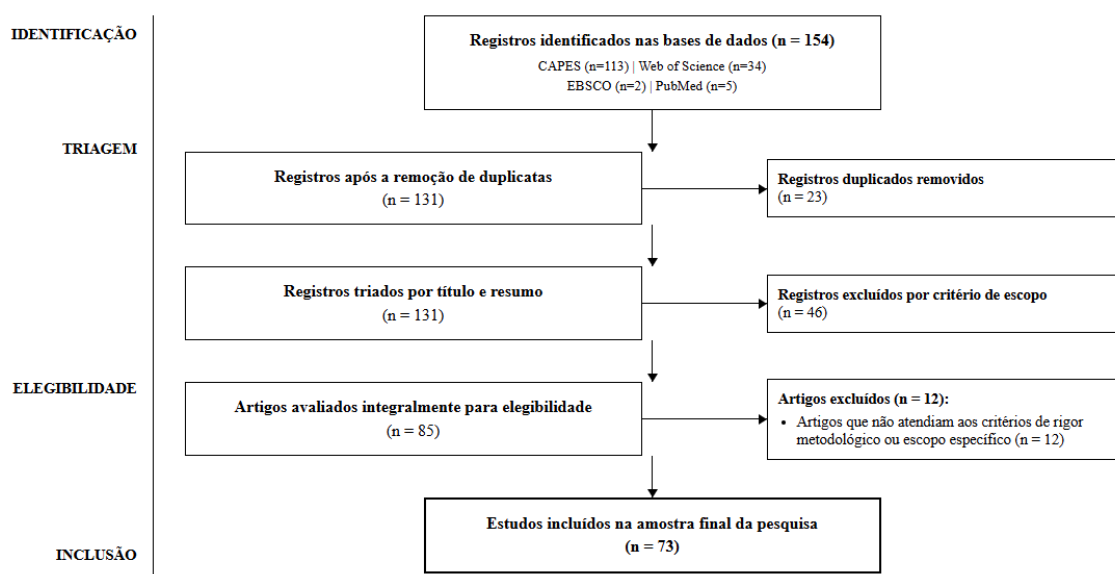
Base de Dados	Filtros e Estratégia de Busca	Resultados Brutos	Duplicatas Removidas	Total para Triagem
Portal CAPES	<i>PERFORMANCE APPRAISAL</i> (título) <i>AND</i> <i>EFFECTIVENESS</i> (título); Filtros: Artigo, Revisado por Pares.	113	-	113
Web of Science	<i>PERFORMANCE APPRAISAL</i> (título) <i>AND</i> <i>EFFECTIVENESS</i> (título).	34	16	18
EBSCO	<i>PERFORMANCE APPRAISAL</i> (título) <i>AND</i> <i>EFFECTIVENESS</i> (título).	2	2	0
PubMed	<i>PERFORMANCE APPRAISAL</i> (título) <i>AND</i> <i>EFFECTIVENESS</i> (título).	5	5	0
TOTAL		154	23	131

Nota. Embora o levantamento bibliográfico tenha abrangido um total de 50 bases de dados e editoras científicas acessadas via portal de periódicos CAPES, a tabela apresenta apenas as fontes que se destacaram pelo maior volume de artigos recuperados.

O processo de seleção seguiu um funil de refinamento focado na profundidade das evidências. Inicialmente, 85 artigos foram pré-selecionados por abordarem a efetividade da AD. Contudo, procedeu-se a um segundo estágio de triagem para garantir a aderência ao objetivo central deste estudo, que busca mapear os antecedentes e resultados do processo. Durante essa leitura, 12 artigos foram descartados por não atenderem ao escopo específico da pesquisa, uma vez que focavam em aspectos gerais da literatura sem detalhar os mecanismos de funcionamento do sistema. Assim, a amostra final foi composta por 73 artigos exclusivos, selecionados por sua relevância direta para a síntese dos determinantes de efetividade.

Figura 1

Fluxograma PRISMA de seleção dos estudos.

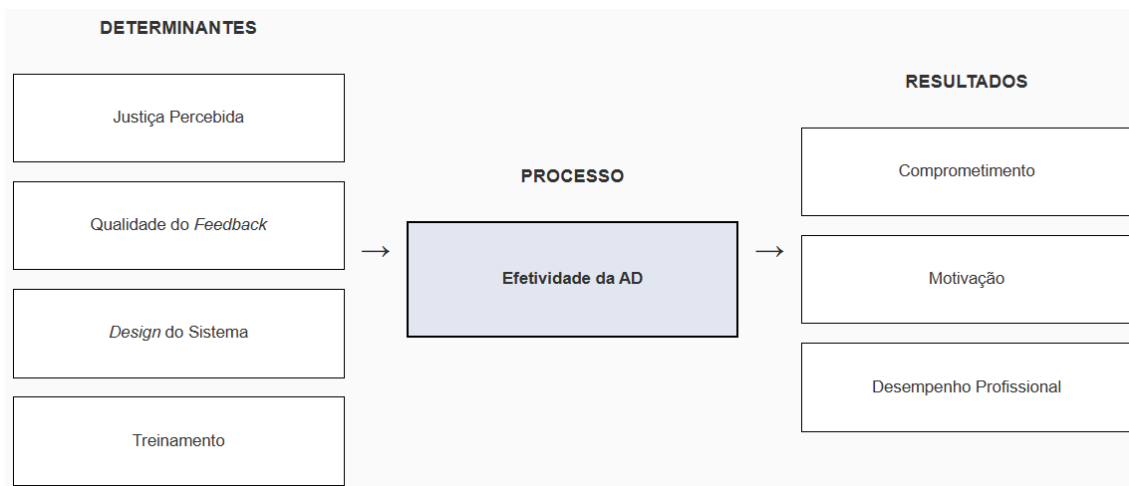


A extração de dados foi conduzida por meio de uma matriz de categorização no Microsoft Excel, em conformidade com o item 10 do protocolo PRISMA (Page et al., 2022). A matriz foi estruturada para coletar dados descritivos (autoria, ano, base de dados) e qualitativos (objetivos e principais achados). Posteriormente, os dados qualitativos foram submetidos a um processo de codificação temática para preenchimento de eixos analíticos pré-definidos na pesquisa: (1) contexto e *design*, focando em variáveis estruturais do sistema (Haines et al., 2004); (2) determinantes e processos, englobando elementos psicossociais como justiça percebida e qualidade do *feedback* (Abdullah & Malik, 2022); e (3) resultados organizacionais, voltados a desfechos como comprometimento e produtividade (Girdharwal, 2019). Esse procedimento permitiu rastrear desde a implementação técnica (Artati et al., 2024) até as percepções subjetivas dos usuários (Benedict & Levine, 1988).

A análise e integração dos dados seguiram a abordagem de síntese narrativa (Galvão & Pereira, 2014). Esta técnica permitiu integrar achados de estudos com diferentes metodologias, desde modelos de treinamento de avaliadores (Davis & Mount, 1984) até análises sobre a efetividade no ensino (Yadav et al., 2020). Os resultados foram agrupados em categorias temáticas que explicam a cadeia causal da AD, sendo seus antecedentes (determinantes) e consequentes (resultados organizacionais).

Figura 2

Modelo Teórico-Conceitual dos Mecanismos de Efetividade da AD.



O presente estudo integra o projeto de pesquisa intitulado “Efetividade das Avaliações de Desempenho nas Organizações”, vinculado ao Instituto de Psicologia da Universidade de Brasília (IP/UnB). O projeto macro foi submetido e aprovado pelo Comitê de Ética em Pesquisa (CEP), estando registrado sob o Parecer nº 8.050.720. A condução deste estudo seguiu rigorosamente as diretrizes éticas vigentes para pesquisas científicas, assegurando a transparência e o compromisso com as normas de integridade acadêmica.

Resultados

No que tange à mensuração da efetividade nos estudos analisados, identificou-se uma pluralidade de abordagens metodológicas. A aferição predominante ocorre por meio de indicadores de percepção, utilizando escalas psicométricas (tipo Likert) validadas para medir a “efetividade percebida” a “utilidade” e a “satisfação com o sistema” sob a ótica de avaliados e avaliadores (Abdullah & Malik, 2022; Modipane et al., 2019). Outra vertente foca na acurácia psicométrica, onde a efetividade é aferida pela redução estatística de erros de julgamento, como o erro de *halo* e o viés de negatividade, frequentemente em desenhos experimentais (Onal & Kumkale, 2022). Por fim, uma terceira abordagem utiliza indicadores indiretos de resultados organizacionais,

em que a efetividade do sistema é inferida a partir de ganhos tangíveis em produtividade, redução de rotatividade e níveis de comprometimento organizacional (Tahsildari & Shahnaei, 2015; Girdharwal, 2019).

Com relação aos determinantes processuais e psicossociais da efetividade, os artigos indicam que a percepção de justiça organizacional é o preditor mais consistente da efetividade. A justiça interacional, que é o modo como o gestor trata o subordinado durante o ciclo avaliativo, demonstrou ter o maior impacto na satisfação e na efetividade percebida (Abdullah & Malik, 2022). Além disso, a clareza dos critérios e a transparência nas políticas de implementação são fundamentais para que o sistema seja aceito como legítimo (Panda et al., 2018).

Já a qualidade do *feedback* e a interação líder-membro (LMX) emergem como mecanismos críticos. A literatura aponta que a habilidade do gerente em conduzir a discussão de feedback é o melhor preditor de uma ação construtiva por parte do trabalhador (Meyer & Walker, 1961). Por outro lado, o processo é frequentemente vulnerável a influências implícitas. Observou-se que os avaliadores tendem a atrasar agendamentos e distorcer notas (especialmente para desempenhos moderadamente baixos) quando o feedback face a face é obrigatório, o que afeta a efetividade real do sistema (Benedict & Levine, 1988).

Tabela 3

Frequência dos Determinantes Críticos de Efetividade.

Determinante Crítico	n	%	Artigos de Referência (Exemplos)
Qualidade do <i>Feedback</i>	28	38,4%	Davis & Mount (1984); Suhasini & Koneru (2016)
Justiça Percebida	22	30,1%	Abdullah & Malik (2022); Modipane et al. (2019)
Clareza de Metas e Padrões	18	24,6%	Nazaruddin et al. (2021); Panda et al. (2018)

Treinamento de Avaliadores	16	21,9%	Davis & Mount (1984); Onal & Kumkale (2022)
Suporte da Liderança / LMX	14	19,2%	Abdulaziz et al (2023); Panda et al. (2018)
Uso de Tecnologia (e-PA)	7	9,6%	Artati et al. (2024); Bensi et al. (2024); Keerthivasan & Meera (2025)

Nota. A soma das porcentagens excede 100% devido à natureza multidimensional dos estudos analisados, pois um mesmo artigo pode identificar e discutir múltiplos determinantes da efetividade, sendo contabilizado em todas as categorias pertinentes dentro da amostra de $n = 73$.

Sobre o treinamento de avaliadores e a mitigação de vieses, a efetividade da AD depende da capacidade dos avaliadores em realizar julgamentos precisos. O treinamento de monitoramento de fontes (*Source-Monitoring Training* - SMT) mostrou-se eficaz em reduzir o viés de negatividade e o erro de *halo*, permitindo que os gestores façam julgamentos baseados em memórias específicas em vez de impressões gerais (Onal & Kumkale, 2022).

De forma complementar, os treinamentos que utilizam instrução assistida por computador (CAI) e modelagem de comportamento aumentam significativamente o conhecimento do avaliador e a qualidade dos planos de desenvolvimento elaborados, resultando em discussões de feedback mais satisfatórias para os subordinados (Davis & Mount, 1984).

Tabela 4

Síntese das Estratégias de Treinamento de Avaliadores e Impacto na Efetividade.

Artigo (Referência)	Tipo de Treinamento	Erro Mitigado	Ganhos em Efetividade Reportados
Onal & Kumkale (2022)	<i>Source-Monitoring Training</i> (SMT)	Viés de Negatividade e Erro de <i>Halo</i> .	Redução significativa do viés de negatividade ($d = 0,56$); aumento na acurácia do reconhecimento comportamental ($d = 0,68$).

Davis & Mount (1984)	<i>Computer-Assisted Instruction (CAI)</i> e Modelagem de Comportamento.	Lacunas de Conhecimento e Habilidade de Comunicação.	Melhora na qualidade dos planos de desenvolvimento; aumento na satisfação dos subordinados com a discussão de <i>feedback</i> .
Panda et al. (2018)	Treinamento em Recompensa e Desenvolvimento.	Falta de objetividade e clareza.	Identificado como fator crucial para a eficácia, contribuindo para explicar 87% da variância total do sistema.
Sánchez Elvira & Bruno-Faria (2021)	Treinamento Gerencial e de Liderança (Foco em <i>Feedback</i>).	Leniência (notas máximas generalizadas).	A ausência de treinamento foi apontada como causa da degradação do sistema em um “ritual burocrático” sem impacto real.
Keerthivasan & Meera (2025)	Desenvolvimento de Competências Avaliativas.	Viés na avaliação e Falta de Transparência.	Treinamento e desenvolvimento explicaram, junto a outros fatores, 57,3% da variância no desempenho dos funcionários.

Nota. Os ganhos reportados referem-se a melhorias estatisticamente significativas ou percepções majoritárias de efetividade nos estudos citados.

Seguindo para o design, tecnologia e adequação metodológica, os estudos sugerem que não existe um “melhor método” ou um “método universal”, mas sim uma necessidade de alinhamento (*fit*) entre o tipo de tarefa e o processo avaliativo. As tarefas que são programáveis se beneficiam de avaliações comportamentais, enquanto tarefas que não são programáveis (focadas em fins) exigem avaliações por resultados para manter a efetividade (Lee, 1985).

No âmbito das ferramentas, o uso de escalas de observação comportamental (BOS) e métodos modernos como o BSC (*Balanced Scorecard*) ou sistemas *online* tem sido adotados para reduzir a subjetividade (Jindal et al., 2015). No entanto, o *design* pode falhar se assumir relações lineares simplistas entre a frequência de comportamentos e a efetividade do desempenho real (Kane & Bernardin, 1982).

A digitalização, por meio de sistemas de *e-performance*, tem melhorado a conformidade e automatizado tarefas administrativas, permitindo maior foco em

atividades estratégicas, embora enfrente barreiras como a falta de habilidades técnicas dos operadores (Artati et al., 2024).

Para os resultados organizacionais e as barreiras contextuais, uma AD dotada de efetividade atua como um catalisador para o comprometimento organizacional. As facetas do processo, como o planejamento da sessão e múltiplos *inputs*, predizem positivamente o compromisso afetivo e de continuidade dos funcionários (Girdharwal, 2019). Além disso, a efetividade percebida reduz a intenção de rotatividade (*turnover*) por meio do cumprimento do contrato psicológico e do aumento do engajamento no trabalho (Sharma et al., 2022).

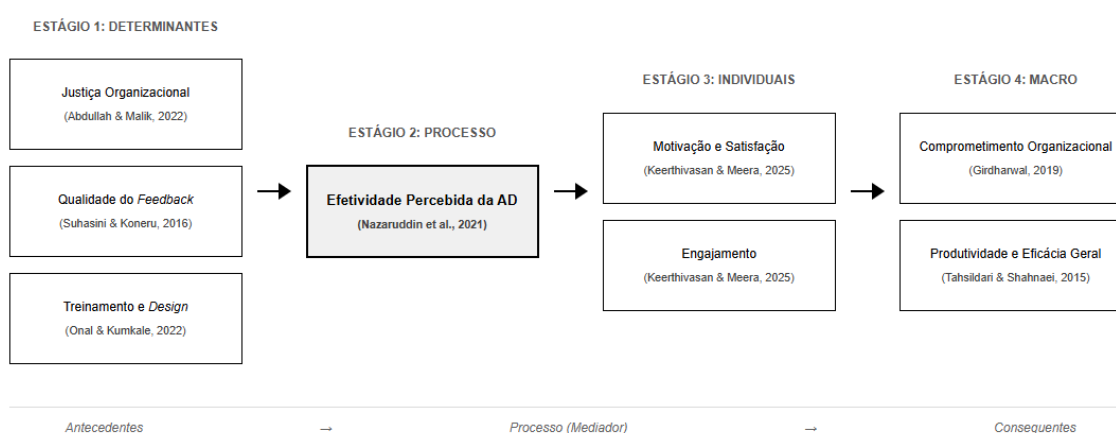
No nível macro, sistemas robustos explicam até 44% da variação na efetividade organizacional geral, impactando positivamente a adaptabilidade, flexibilidade e produtividade das instituições (Sahu et al., 2014). Contudo, em contextos de setor público e economias emergentes, a efetividade é frequentemente ameaçada por barreiras políticas e culturais. A politização, a corrupção e o “coleguismo” podem transformar a AD em um rito burocrático e ineficaz na prática, independentemente de quão sofisticado seja o sistema (Sakpere, 2022; Sánchez Elvira & Bruno-Faria, 2021).

A partir da análise integrativa dos determinantes, processos e desfechos reportados nos 73 estudos, foi elaborada uma síntese gráfica para ilustrar a cadeia causal da efetividade da AD. A Figura 3 apresenta este modelo teórico-conceitual proposto, que organiza os achados em um fluxo lógico. O modelo postula que a efetividade não é um evento isolado, mas um processo mediador central. Conforme ilustrado, os determinantes técnicos e humanos (como a justiça e o *feedback*) não impactam diretamente os resultados finais, eles primeiramente constroem a “efetividade percebida” do sistema. É esta percepção de utilidade e justiça que, por sua vez,

desencadeia as atitudes individuais positivas (motivação, satisfação) que culminam em ganhos organizacionais tangíveis (produtividade, comprometimento), conforme evidenciado por estudos como os de Keerthivasan & Meera (2025) e Tahsildari & Shahnaei (2015).

Figura 3

Modelo de Impacto da Efetividade da AD.



Nota. O modelo representa uma síntese conceitual elaborada com base na integração dos achados da revisão sistemática ($n = 73$). A efetividade percebida é posicionada como a variável mediadora central que converte os insumos (determinantes) em resultados comportamentais e operacionais.

Discussão

Os achados desta síntese consolidam uma mudança de paradigma na literatura de gestão de pessoas: o deslocamento do foco da “precisão psicométrica” para a “utilidade social”. Enquanto os estudos pioneiros da década de 80 e 90 se concentraram quase exclusivamente na arquitetura de escalas para mitigar erros de julgamento, as evidências atuais revelam que a AD opera como um sistema aberto, sensível ao

contexto político e às trocas sociais. Essa transição sugere que a efetividade não é uma propriedade intrínseca ao formulário de avaliação, mas um fenômeno emergente da qualidade do diálogo entre líder e liderado.

Um dos achados desta síntese é que as facetas do processo (como o planejamento da sessão e a qualidade da troca líder-membro) são preditores mais fortes da efetividade do que as facetas do sistema (o design do formulário em si). A evidência de que a justiça interacional possui o maior impacto na satisfação e na efetividade percebida (Abdullah & Malik, 2022) reforça a tese de que o “como” se avalia é mais crítico do que o “quê” se avalia.

Nesse sentido, o modelo de “cadeia de reações” proposto por Maharvi et al. (2023) oferece uma explicação processual para a efetividade. Segundo os autores, a qualidade da troca entre líder e liderado (LMX) funciona como o antecedente primário, um suporte de liderança robusto sinaliza ao funcionário que o processo avaliativo é pautado pela benevolência e integridade. Esta sinalização desencadeia uma sequência de reações cognitivas e afetivas, primeiro, o indivíduo valida a justiça procedimental (o método é justo) e a justiça interacional (o tratamento é digno), essa validação gera, conseqüentemente, a aceitação do *feedback* e a satisfação com o sistema.

A “efetividade”, portanto, não é um atributo estático do sistema, mas o resultado final dessa cadeia de reações positivas. Esse modelo ajuda a explicar por que métodos tecnicamente sofisticados, como as Escalas de Observação Comportamental (BOS), frequentemente falham na prática. Conforme criticado por Kane & Bernardin (1982), escalas como a BOS se baseiam na premissa linear de que a simples observação e contagem da frequência de comportamentos garantiriam a acurácia. Contudo, o modelo de Maharvi et al. (2023) demonstra que, sem a “base de confiança” inicial da liderança,

o trabalhador tende a reagir com ceticismo à métrica, independentemente de sua precisão técnica, rompendo a cadeia que levaria à efetividade organizacional.

Apesar do avanço nos treinamentos de acurácia, os artigos revelam uma lacuna persistente entre a intenção do *design* e a execução prática. O fenômeno do “atraso e distorção” (Benedict & Levine, 1988) exemplifica como as pressões interpessoais levam os avaliadores a inflar notas para evitar o desconforto do *feedback* negativo ou mais focado no desenvolvimento, drenando a efetividade estratégica da AD.

Esse cenário é agravado em contextos de setor público, onde a estabilidade e a cultura de “coleguismo” frequentemente transformam a AD em um rito burocrático e leniente (Sánchez Elvira & Bruno-Faria, 2021). A politização do processo emerge, portanto, como a principal barreira à efetividade, sugerindo que o sistema só gera valor real quando consegue se desvincular de interesses pessoais e corrupção institucional (Sakpere, 2022).

As análises dos estudos confirmam que a efetividade da AD atua como uma “âncora” para o contrato psicológico entre trabalhador e organização. O impacto positivo no comprometimento afetivo (Girdharwal, 2019) demonstra que, quando o funcionário percebe a avaliação como uma ferramenta de desenvolvimento e não apenas de punição, sua ligação emocional com a instituição é fortalecida.

A integração da AD com práticas de treinamento e participação, observada nesta revisão, corrobora a evolução da gestão de recursos humanos de uma perspectiva isolada para uma abordagem de “pacotes de práticas” (*HR bundles*). Enquanto estudos das décadas de 1990 e 2000 focavam na AD como uma ferramenta administrativa autônoma, as evidências atuais de Tahsildari & Shahnaei (2015) reforçam que o impacto na efetividade organizacional macro é sinérgico. Comparativamente a revisões clássicas,

percebe-se que a tecnologia (SIRH) deixou de ser apenas um suporte de armazenamento para se tornar um vetor de transparência e democratização de dados (Bensi et al., 2024; Khashman & Khashman, 2016). Essa evolução aponta que a efetividade não decorre mais da “vigilância” eletrônica, mas da acessibilidade à informação, o que representa um salto qualitativo em relação aos sistemas puramente burocráticos do passado.

Não obstante os avanços identificados, a literatura atual ainda enfrenta entraves metodológicos críticos, com destaque para a predominância de estudos transversais. Essa limitação persistente impede a verificação da estabilidade da efetividade ao longo do tempo, deixando lacunas sobre se um sistema percebido como justo em um ciclo específico mantém seu impacto motivacional após sucessivas janelas de bonificação ou em cenários de cortes orçamentários. Somado a isso, a dependência quase exclusiva de medidas de autorrelato eleva o risco de viés de variância comum, sugerindo que parte das correlações reportadas entre justiça e satisfação pode estar inflada pela subjetividade dos respondentes.

Diante dessas lacunas, a agenda de pesquisa futura deve priorizar investigações sobre a longevidade do sistema, analisando como a percepção de efetividade evolui ao longo de carreiras de longo prazo por meio de delineamentos longitudinais. Igualmente, buscar a convergência metodológica ao analisar o grau de discrepância entre as medidas subjetivas de satisfação e indicadores objetivos de produtividade, como o alcance de metas e a redução de erros técnicos.

De forma adicional, as transformações tecnológicas impõem a necessidade de compreender de que maneira a introdução de algoritmos e inteligência artificial na análise de desempenho altera a percepção de justiça interacional e a natureza do *feedback* humano. Por fim, recomenda-se que futuros estudos adotem perspectivas

multinível, capazes de desvendar como as subculturas das unidades de negócio moderam a relação entre o *design* organizacional do sistema e a motivação individual, permitindo uma compreensão mais holística e contextualizada da efetividade da AD.

Conclusão

O presente estudo permitiu a consolidação de um modelo compreensivo sobre a efetividade da Avaliação de Desempenho (AD), evidenciando que o sucesso deste sistema não reside na sofisticação tecnológica de seus formulários, mas na qualidade das relações humanas e na percepção de equidade que o sustentam.

Contribuições Teóricas

A principal contribuição deste estudo reside na mudança de paradigma proposta: a transição do foco tradicional na precisão psicométrica das escalas para a compreensão da AD como um processo psicossocial mediado. O estudo inova ao sintetizar uma cadeia causal clara onde a efetividade percebida atua como o elo central que transforma insumos técnicos (*design* e treinamento) em resultados comportamentais (motivação) e estratégicos (produtividade).

Os achados corroboram a literatura clássica e contemporânea ao reafirmar que o “fator humano” traduzido pela qualidade do *feedback* e pela confiança na relação líder-membro (LMX), é o principal propulsor de aceitação do sistema. A evidência de que a justiça interacional e procedimental são os preditores mais consistentes de satisfação alinha-se às teorias de justiça organizacional, demonstrando que o modo como o processo é conduzido é mais crítico para a efetividade do que o método de avaliação em si. O estudo ainda aponta sobre a efetividade de treinamentos cognitivos modernos, como o monitoramento de fontes (SMT), como substitutos superiores aos treinamentos de erros de escala tradicionais.

Implicações Práticas

Do ponto de vista aplicado, o estudo contribui para a gestão de Recursos Humanos ao sugerir que as organizações devem priorizar o desenvolvimento de competências comunicativas na liderança e a transparência processual em detrimento da busca por “métodos infalíveis”. A integração de tecnologias digitais e análise de dados é apontada como uma via para reduzir a opacidade e a politização do sistema, especialmente em contextos onde o “coleguismo” e a leniência comprometem a utilidade da avaliação.

Limitações do Estudo

Apesar do rigor na seleção dos 73 artigos, esta revisão apresenta limitações que devem ser consideradas. Primeiramente, a predominância de delineamentos transversais na literatura limita a capacidade de inferir causalidade definitiva sobre a sustentabilidade da efetividade ao longo de ciclos anuais sucessivos. Em segundo lugar, a forte dependência de medidas de autorrelato nos estudos primários pode introduzir o viés de variância comum, onde a satisfação com o sistema pode inflar a percepção de sua efetividade real. Por fim, observou-se uma concentração de estudos em setores específicos, o que pode dificultar a generalização integral para contextos de economias emergentes com culturas organizacionais altamente politizadas.

Agenda de Pesquisa

Para o avanço da área, sugere-se que pesquisas futuras adotem desenhos longitudinais para monitorar como a percepção de efetividade evolui após mudanças estruturais no *design* da AD. Recomenda-se também o cruzamento de percepções

subjetivas com dados objetivos de produtividade e indicadores de rotatividade (*turnover*) para validar o impacto econômico do sistema.

No campo tecnológico, pode-se investigar como o uso de inteligência artificial e análise de sentimentos no *feedback* pode aumentar a percepção de transparência ou, de forma alternativa, gerar resistências por desumanizar o processo. Por fim, encoraja-se a exploração de como variáveis culturais (coletivismo e individualismo) moderam a relação entre a justiça percebida e o compromisso afetivo, garantindo que os modelos de efetividade sejam aplicáveis a um cenário globalizado e diverso.

Referências

- Abdulaziz, T., Som, H. M., Mahamood, R., & Nadarajah, D. (2023). Leadership development interventions, job satisfaction, and leadership effectiveness as determinants of organisational performance in Saudi transportation sector. *International Journal of Academic Research in Business and Social Sciences*, 13(10), 1805–1824. <https://doi.org/10.6007/IJARBSS/v13-i10/18977>
- Abdullah, Z., & Malik, S. K. (2022). Perceived appraisal fairness effect on performance appraisal effectiveness and appraisal satisfaction in higher education. *Journal of Business and Social Review in Emerging Economies*, 8(2), 287–300. <https://doi.org/10.26710/jbsee.v8i2.2237>
- Artati, A., Nasir, M., & Nursani, N. (2024). Analysis of the effectiveness of the implementation of e-performance at the Office of the Department of Trade and Industry of Gowa Regency. *Eduvest - Journal of Universal Studies*, 4(10), 8765–8777. <https://doi.org/10.59188/eduvest.v4i10.34665>
- Benedict, M. E., & Levine, E. L. (1988). Delay and distortion: Tacit influences on performance appraisal effectiveness. *Journal of Applied Psychology*, 73(3), 507–514. <https://doi.org/10.1037/0021-9010.73.3.507>
- Bensi, M. E., Bensi, L. P., Oliveros, A. G. G., Bantug, E. T., & Pula, R. L. (2024). Enhancing faculty performance appraisals with sentiment analysis and data visualization for evaluating teaching effectiveness. *International Journal of Research and Scientific Innovation*, 11(10), 577–592. <https://doi.org/10.51244/IJRSI.2024.1110049>
- 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>

Cascio, W. F. (2019). *Managing human resources: Productivity, quality of work life, profits* (11.^a ed., pp. 340–345). McGraw Hill Education.

Davis, B. L., & Mount, M. K. (1984). Effectiveness of performance appraisal training using computer assisted instruction and behavior modeling. *Personnel Psychology*, 37(3), 439–452.

<https://doi.org/10.1111/j.1744-6570.1984.tb00521.x>

Galvão, T. F., & Pereira, M. G. (2014). Revisões sistemáticas da literatura: passos para sua elaboração. *Epidemiologia e Serviços de Saúde*, 23(1), 183–184.

<https://doi.org/10.5123/S1679-49742014000100018>

Girdharwal, N. (2019). A comparative study of performance appraisal on effectiveness and organizational commitment. *Prabandhan: Indian Journal of Management*, 12(11), 45–53. <https://doi.org/10.17010/pijom/2019/v12i11/148409>

Günther, H. (2004). *Como elaborar um relato de pesquisa* (Série: Planejamento de Pesquisa nas Ciências Sociais, No. 02). Laboratório de Psicologia Ambiental, Universidade de Brasília. <http://www.unb.br/ip/lpa/pdf/02Sugestoes.pdf>

Haines, V. Y., III, St-Onge, S., & Marcoux, A. (2004). Performance management design and effectiveness in quality-driven organizations. *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, 21(2), 146–161. <https://doi.org/10.1111/j.1936-4490.2004.tb00330.x>

Jindal, S., Laveena, & Navneet. (2015). Study of effectiveness of performance appraisal system in selected Indian companies. *International Journal of Research -*

GRANTHAALAYAH, 3(12), 44–50.

<https://doi.org/10.29121/granthaalayah.v3.i12.2015.2885>

Kane, J. S., & Bernardin, H. J. (1982). Behavioral observation scales and the evaluation of performance appraisal effectiveness. *Personnel Psychology*, 35(3), 635–641.

<https://doi.org/10.1111/j.1744-6570.1982.tb02215.x>

Keerthivasan, S., & Meera, A. (2025). An analytical study on the effectiveness of performance appraisal systems in enhancing employee motivation and organizational productivity. *International Journal of Scientific Research in Engineering and Management*, 9(4), 1–9.

<https://doi.org/10.55041/IJSREM45703>

Khashman, I., & Khashman, A. (2016). The impact of human resource information system (HRIS) applications on organizational performance (efficiency and effectiveness) in Jordanian private hospitals. *Journal of Management Research*, 8(3), 31–51. <https://doi.org/10.5296/jmr.v8i3.9419>

Lee, C. (1985). Increasing performance appraisal effectiveness: Matching task types, appraisal process, and rater training. *The Academy of Management Review*, 10(2), 322–331. <https://doi.org/10.2307/257973>

Maharvi, M. W., Kumar, A., Channa, K. A., & Mahmood, A. (2023). How leader member exchange affects effectiveness of performance appraisal system: A chain of reactions model. *Cogent Business & Management*, 10(2).

<https://doi.org/10.1080/23311975.2023.2238392>

Martin, D. C., & Bartol, K. M. (1998). Performance appraisal: Maintaining system effectiveness. *Public Personnel Management*, 27(2), 223–230.

<https://doi.org/10.1177/009102609802700208>

Meyer, H. H., & Walker, W. B. (1961). A study of factors relating to the effectiveness of a performance appraisal program. *Personnel Psychology*, *14*(4), 291–298.

<https://doi.org/10.1111/j.1744-6570.1961.tb01235.x>

Modipane, P., Botha, P., & Blom, T. (2019). Employees' perceived effectiveness of the performance management system at a North-West provincial government department. *SA Journal of Human Resource Management*, *17*, 12 pages.

<https://doi.org/10.4102/sajhrm.v17i0.1081>

Nazaruddin, I., Sofyani, H., & Zailan, R. Z. (2022). Implications of strategic and development purposes on the effectiveness of performance appraisals and motivation to improve. In *Proceedings of the International Conference on Sustainable Innovation Track Accounting and Management Sciences (ICOSIAMS 2021)* (pp. 193–196). Atlantis Press.

<https://doi.org/10.2991/aebmr.k.211225.027>

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, *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. (2022). A declaração PRISMA 2020: diretriz atualizada para relatar revisões sistemáticas. *Revista Panamericana de Salud Pública*, *46*, e112.

<https://doi.org/10.26633/RPSP.2022.112>

- Panda, S., Waris, M., Asadullah, K., Mehfooz, U., & Adeleke, A. Q. (2018). Analyzing factors attributing to effectiveness of performance management system of a manufacturing industry. *The Journal of Social Sciences Research*, (Special Issue 6), 907–914. <https://doi.org/10.32861/jssr.spi6.907.914>
- Roberts, G. E. (1992). Linkages between performance appraisal system effectiveness and rater and ratee acceptance. *Review of Public Personnel Administration*, 12(3), 19–41. <https://doi.org/10.1177/0734371X9201200303>
- Sánchez Elvira, E. J., & Bruno-Faria, M. de F. (2021). Impact of policy implementation shortcomings and context constraints on the effectiveness of civil servant performance appraisal: the case of Brazilian Federal Service. *Revista Do Serviço Público*, 72(b), 88 - 115. <https://doi.org/10.21874/rsp.v72.ib.5941>
- Sahu, R. N., Jena, L. K., & Parida, S. C. (2014). Performance management system as a predictor of organizational effectiveness: Insights from Indian manufacturing industries. *Jindal Journal of Business Research*, 3(1-2), 137–152. <https://doi.org/10.1177/2278682115627214>
- Sakpere, R. (2022). Effectiveness of employees' performance appraisal in the Nigerian public sector: An empirical study. *World Journal of Social Science Research*, 8(3), 27–41. <https://doi.org/10.22158/wjssr.v8n3p27>
- Sharma, N. P., Sharma, T., & Agarwal, M. N. (2022). Relationship between perceived performance management system (PMS) effectiveness, work engagement and turnover intention: Mediation by psychological contract fulfillment. *Benchmarking: An International Journal*, 29(9), 2985–3007. <https://doi.org/10.1108/BIJ-01-2021-0008>
- Suhasini, A., & Koneru, K. (2016). A study on effectiveness of performance appraisal system in educational institutions. *SSRG International Journal of Economics*

and Management Studies, 3(11), 22–24.

<https://doi.org/10.14445/23939125/IJEMS-V3I6P111>

Tahsildari, A., & Shahnaei, S. (2015). Enhancing organizational effectiveness by performance appraisal, training, employee participation, and job definition.

European Journal of Business and Management, 7(12), 56–63.

<https://iiste.org/Journals/index.php/EJBM/article/view/21573>

Yadav, K. P., Sharma, K., & Dixit, S. (2020). Role of performance appraisal system in enhancing teaching effectiveness: A study of school teachers. *Pacific Business Review International*, 13(1), 23–32.

http://www.pbr.co.in/2020/2020_month/July/3.pdf

**Mensurando a Efetividade da Avaliação de Desempenho:
Desenvolvimento de uma Escala para o Contexto Organizacional.**

Measuring the Effectiveness of Performance Appraisal:

Developing a Scale for the Organizational Context.

Resumo

A Avaliação de Desempenho (AD) é uma prática estratégica essencial. No entanto, sua efetividade, compreendida como a percepção de utilidade, justiça e alcance real dos objetivos de desenvolvimento e gestão, é frequentemente limitada por percepções de burocracia e falta de sentido prático. Este estudo teve como objetivo desenvolver e buscar evidências de validade para a Escala de Efetividade da Avaliação de Desempenho (EEAD) no contexto organizacional brasileiro. Adotou-se um delineamento de métodos mistos, iniciando com uma etapa qualitativa (grupos focais) para definição do construto e geração de itens. A validade de conteúdo foi atestada por seis juízes especialistas (CVC geral = 0,68). Procedeu-se à etapa quantitativa com duas amostras independentes. A Análise Fatorial Exploratória (AFE; $n = 483$) indicou uma estrutura unidimensional explicando 65,1% da variância total, resultando em um instrumento refinado de 31 itens com cargas fatoriais superiores a 0,77. A Análise Fatorial Confirmatória (AFC; $n = 346$) corroborou o modelo com excelentes índices de ajuste: $CFI = 0,993$; $TLI = 0,992$; $RMSEA = 0,064$. O instrumento demonstrou precisão elevada ($\omega = 0,986$) e estabilidade do fator (Índice H = 0,988). Adicionalmente, obteve-se evidência de validade convergente através da correlação positiva com a dimensão “crenças” ($p = 0,67$). Conclui-se que a EEAD é um instrumento robusto e fidedigno, capaz de subsidiar uma gestão baseada em evidências ao diagnosticar a efetividade percebida dos sistemas de AD nas organizações.

Palavras-chave: Avaliação de Desempenho; Efetividade; Construção de Escalas; Psicometria; Validade.

Abstract

Performance Appraisal (PA) is an essential strategic practice. However, its effectiveness, understood as the perception of usefulness, fairness, and real achievement of development and management objectives, is often limited by perceptions of bureaucracy and a lack of practicality. This study aimed to develop and seek validity evidence for the Performance Appraisal Effectiveness Scale (PAES) in the Brazilian organizational context. A mixed-methods design was adopted, beginning with a qualitative stage (focus groups) to define the construct and generate items. Content validity was attested by six expert judges (overall CVC = 0.68). A quantitative stage was then conducted with two independent samples. Exploratory Factor Analysis (EFA; N = 483) indicated a unidimensional structure explaining 65.1% of the total variance, resulting in a refined instrument of 31 items with factor loadings greater than 0.77. Confirmatory Factor Analysis (CFA; N = 346) corroborated the model with excellent fit indices: CFI = 0.993; TLI = 0.992; RMSEA = 0.064. The instrument demonstrated high precision (McDonald's Omega = 0.986) and factor stability (H-index = 0.988). Additionally, evidence of convergent validity was obtained through the positive correlation with the "beliefs" dimension ($p = 0.67$). It is concluded that the EEAD is a robust and reliable instrument, capable of supporting evidence-based management by diagnosing the perceived effectiveness of performance appraisal systems in organizations.

Keywords: Performance Appraisal; Effectiveness; Scale Construction; Psychometrics; Validity.

Introdução

A Avaliação de Desempenho (AD) é definida como o processo sistemático de mensuração e análise do comportamento profissional e dos resultados alcançados pelos trabalhadores em relação aos objetivos organizacionais (Longo, 2007), amplamente reconhecida como uma das práticas mais sofisticadas e, simultaneamente, mais problemáticas da gestão de pessoas. Embora seja considerada essencial para o alinhamento estratégico e o desenvolvimento de competências, o investimento organizacional na implementação desses sistemas muitas vezes não se traduz em melhorias tangíveis na produtividade ou no clima organizacional. Frequentemente, a AD é reduzida a um ritual anual de preenchimento de formulários, sendo percebida por gestores e funcionários como um processo meramente burocrático e desconectado da realidade laboral (Girdharwal, 2019; Sharma et al., 2022).

A efetividade é compreendida como um construto psicossocial que depende da aceitação, da percepção de utilidade e da qualidade das interações geradas pelo processo (Findley et al., 2000; Sharma et al., 2022). O desafio reside no fato de que a funcionalidade técnica de um sistema de AD não garante sua efetividade. Na literatura internacional, diferentes abordagens foram adotadas para tentar operacionalizar esse construto. Findley et al. (2000), por exemplo, propuseram que a efetividade deve ser mensurada a partir da distinção entre as facetas do “sistema” (regras e políticas formais) e as facetas do “processo” (a qualidade da interação e do planejamento da sessão entre líder e liderado). Por outro lado, Ikramullah et al. (2016) buscaram uma visão multidimensional baseada no *Competing Values Framework* (CVF), sugerindo que a efetividade da AD deve ser avaliada sob quatro perspectivas: processos internos, relações humanas, sistema aberto e objetivos racionais.

Outros estudos focaram em componentes específicos do sistema. Iqbal et al. (2019) integraram a efetividade à análise dos propósitos da avaliação e às reações dos avaliados, utilizando instrumentos que medem o quanto o sistema é percebido como útil para o desenvolvimento versus para fins administrativos. Já estudos como os de Modipane et al. (2019), baseados nas escalas de Sharma et al. (2016), concentraram a mensuração na “precisão” percebida do sistema e na sua relação com a justiça organizacional. As abordagens clássicas como as de Longenecker et al. (1988) já destacavam que a efetividade percebida depende de itens que capturam a clareza dos padrões e a utilidade prática do *feedback* recebido.

No entanto, nota-se que muitas dessas métricas são fragmentadas, focando na precisão, na justiça, ou foram desenvolvidas para contextos culturais muito distintos do realismo organizacional brasileiro. Apesar dessa diversidade de esforços, ainda se observa uma escassez de instrumentos psicométricos com indícios de validade que permitam mensurar a efetividade da AD de forma integrada e unidimensional no país. A ausência de uma ferramenta científica nacional impede que organizações realizem auditorias precisas de seus processos, resultando em gestões baseadas em intuições e não em evidências

A necessidade de criação da Escala de Efetividade da Avaliação de Desempenho (EEAD) justifica-se pela lacuna diagnóstica entre a intenção do RH e a percepção do trabalhador. Conforme evidenciado em grupos focais realizados previamente a este estudo, os profissionais identificaram a necessidade de que a AD atue como um grande diagnóstico institucional, mas relatam que a falta de objetividade e de rituais pós-avaliação esvazia o valor da ferramenta. Diferente de escalas que medem apenas a satisfação com o feedback ou a precisão da nota, a EEAD propõe-se a capturar a

efetividade como um fenômeno unidimensional e holístico, abarcando desde o planejamento da sessão até o impacto estratégico percebido.

A relevância de um instrumento desta natureza é amplificada pelo seu papel como preditor de resultados organizacionais. Alguns estudos indicam que a percepção de um sistema de avaliação efetivo correlaciona-se positivamente com o engajamento no trabalho e negativamente com a intenção de rotatividade (Sharma et al., 2022). Além disso, a validade de um sistema de AD está intrinsecamente ligada ao sistema de crenças dos empregados sobre a justiça e o suporte organizacional (Gonçalves & Leite, 2005). Assim, a EEAD surge como uma solução para transformar a AD de um custo operacional em um ativo estratégico de inteligência de dados.

Diante deste cenário, o presente estudo descreve o processo de desenvolvimento e a busca de indícios de validade da Escala de Efetividade da Avaliação de Desempenho (EEAD). Por meio de um design de métodos mistos, este trabalho integra a análise qualitativa de conteúdo com o rigor estatístico das análises fatoriais exploratórias e confirmatórias. O objetivo é oferecer à comunidade acadêmica e aos profissionais de mercado um instrumento fidedigno, capaz de diagnosticar a efetividade dos sistemas de gestão de desempenho e orientar intervenções que garantam a utilidade real dessa prática nas organizações brasileiras.

Método

O presente estudo caracteriza-se como uma pesquisa de métodos mistos, de caráter exploratório e confirmatório, voltada ao desenvolvimento e análise dos índices de validade de uma escala psicométrica. O procedimento seguiu as etapas recomendadas pela literatura para a construção de instrumentos de medida em Psicologia (Pasquali, 2010), se dividindo em: (1) procedimentos teóricos e qualitativos; (2) procedimentos

semânticos e de conteúdo; e (3) procedimentos estatísticos (validade de estrutura interna).

Procedimentos Teóricos e Qualitativos

Para a definição do construto e geração do banco de itens, foram realizados grupos focais com profissionais que possuíam experiência direta com sistemas de Avaliação de Desempenho (AD). O critério de inclusão estabeleceu que os participantes deveriam ter atuado em pelo menos uma das seguintes frentes: como avaliados, como avaliadores ou na criação e gestão de sistemas de AD em contextos organizacionais.

As sessões foram conduzidas em ambiente virtual, via plataforma *Google Meet* de videoconferência, devido à dispersão geográfica e disponibilidade dos participantes. Foram realizadas 5 sessões com duração entre 60 e 120 minutos cada, ao final da quinta sessão, observou-se a saturação teórica dos temas, não sendo necessária a realização de encontros adicionais. O tamanho dos grupos variou entre 6 e 10 participantes, e as coletas foram encerradas ao se atingir o critério de saturação de respostas. Utilizou-se um roteiro de entrevista semiestruturado composto por questões norteadoras. O objetivo deste roteiro foi estimular a reflexão e instigar a discussão espontânea entre os participantes a partir de suas experiências vividas. As perguntas funcionaram como disparadores para explorar o fenômeno de forma ampla, garantindo que o fluxo da conversa cobrisse a realidade prática do processo de avaliação nas organizações.

O presente estudo integra o projeto de pesquisa intitulado “Efetividade das Avaliações de Desempenho nas Organizações”, vinculado ao Instituto de Psicologia da Universidade de Brasília (IP/UnB). O projeto macro foi submetido e aprovado pelo Comitê de Ética em Pesquisa (CEP), estando registrado sob o Parecer nº 8.050.720. A condução deste estudo seguiu rigorosamente as diretrizes éticas vigentes para pesquisas

científicas, assegurando a transparência e o compromisso com as normas de integridade acadêmica. O Termo de Consentimento Livre e Esclarecido (TCLE) foi disponibilizado digitalmente para leitura e assinatura e o termo foi lido em voz alta no início de cada gravação, assegurando o direito de retirada a qualquer momento.

As gravações foram transcritas integralmente para posterior análise de conteúdo (Bardin, 2016). As transcrições das sessões dos grupos focais foram analisadas em sua totalidade, seguindo o critério de exaustividade. Inicialmente, realizou-se uma leitura flutuante de todo o material para identificar a variedade de percepções expressas. Contudo, a fase de codificação e o posterior agrupamento em temas focaram especificamente nos núcleos de sentido que descreviam critérios, comportamentos e resultados associados à efetividade (ou falta dela) nos processos de AD. Assim, os destaques das falas submetidos ao grupo de pesquisa (13 integrantes) compreenderam todos os elementos do discurso que ajudavam a operacionalizar o construto “efetividade”. Esse processo de busca por consenso resultou na sistematização de oito categorias analíticas, que serviram de base para a redação dos itens, assegurando que o instrumento cobrisse as diversas facetas do fenômeno manifestadas pelos participantes.

Construção do Instrumento e Validação de Conteúdo

A partir das categorias identificadas, procedeu-se à redação dos itens seguindo as diretrizes de Pasquali (2010) para a construção de escalas como clareza, simplicidade, objetividade, tipicidade e credibilidade. Os itens foram redigidos em primeira pessoa, evitando termos negativos para reduzir vieses de resposta. Foram formulados nove itens para cada uma das oito categorias, totalizando um banco inicial de 72 itens, os quais passaram por revisão técnica do grupo de pesquisa antes da etapa de juízes.

A validade de conteúdo foi verificada por um comitê de seis juízes, todos doutores em Psicologia com experiência em gestão de pessoas e psicometria. Os juízes avaliaram os itens quanto aos critérios de clareza de linguagem, pertinência prática e relevância teórica, utilizando uma escala do tipo Likert de cinco pontos, na qual 1 indicava pouquíssima aderência e 5 muitíssima aderência dos itens ao conceito avaliado. Para a análise quantitativa, utilizou-se o Coeficiente de Validade de Conteúdo (CVC), conforme proposto por Hernández-Nieto (2002) e detalhado por Cassepp-Borges et al. (2010).

Após a etapa de juízes, a versão experimental foi submetida a uma análise semântica com uma amostra do público-alvo ($n = 20$). Os resultados dessa etapa evidenciaram a necessidade de ajustes na redação, resultando em mudanças estruturais em determinados itens para eliminar ambiguidades e aprimorar a compreensão, consolidando a versão adequada para a aplicação em larga escala.

Análise Fatorial Exploratória - AFE

Amostra

Participaram dessa etapa profissionais com experiência prévia em sistemas de Avaliação de Desempenho (AD), convidados por meio de redes profissionais e contatos organizacionais. O instrumento de pesquisa e os itens demográficos foram cadastrados na plataforma de pesquisa online *Sosci Survey*. Um link de acesso foi disponibilizado aos participantes para que respondessem de forma individual, voluntária e anônima.

Procedimentos de coleta

Ao acessarem a pesquisa, os respondentes liam um texto com informações sobre os objetivos do estudo, acompanhado de instruções que garantiam o anonimato e a

confidencialidade das informações. Para prosseguir, foi solicitado obrigatoriamente o aceite ao Termo de Consentimento Livre e Esclarecido (TCLE).

Instrumento

Além dos itens da versão experimental da Escala de Efetividade da Avaliação de Desempenho (EEAD), também foram solicitados cinco itens demográficos para caracterização da amostra: tipo de organização trabalhada, gênero, idade, escolaridade e tempo na organização atual.

Análise de Dados

Inicialmente, procedeu-se à limpeza do banco de dados. Foi constatado que nenhum participante precisou ser retirado da amostra, uma vez que a frequência de itens pendentes de resposta foi inexpressiva e não comprometeu a integridade dos dados finais. Os valores omissos identificados foram tratados por meio do método *Hot-Deck Multiple Imputation* (Lorenzo-Seva & Van Ginkel, 2016).

As análises foram implementadas no programa *Factor*, versão 12.06.07 (Ferrando & Lorenzo-Seva, 2017), utilizando uma matriz de correlação policórica, recomendada para variáveis ordinais, e o método de extração *Robust Diagonally Weighted Least Squares* (RDWLS). A decisão sobre o número de fatores a ser retido baseou-se na técnica de Análise Paralela (PA) com permutação aleatória dos dados observados (Timmerman & Lorenzo-Seva, 2011), que sugeriu a retenção de um único fator. A Tabela 4 apresenta a variância explicada com base nos autovalores e os resultados da análise paralela.

Análise Fatorial Confirmatória - AFC

Amostra

Participaram dessa etapa profissionais de distintas empresas atuantes em organizações públicas e privadas. O instrumento de pesquisa foi cadastrado na plataforma de pesquisa online *Sosci Survey* e o link de acesso foi divulgado em redes sociais e enviado por e-mail para resposta individual, voluntária e anônima.

Procedimentos de coleta

Ao acessarem a pesquisa, os respondentes liam um texto com informações sobre o objetivo do estudo, garantia de anonimato e confidencialidade, sendo solicitado o aceite do Termo de Consentimento Livre e Esclarecido (TCLE).

Instrumento

Foram aplicados os 31 itens da Escala de Efetividade da Avaliação de Desempenho (EEAD), resultantes da etapa exploratória, além da dimensão “crenças” do instrumento de Gonçalves e Leite (2005) para fins de validade convergente. Também foram solicitados cinco itens demográficos para caracterização da amostra: tipo de organização, gênero, idade, escolaridade e tempo na organização.

Análise de Dados

Procedeu-se inicialmente à limpeza do banco de dados, sendo identificados e removidos 32 participantes que preencheram exclusivamente as informações sociodemográficas e profissionais, sem responder aos itens do instrumento. Após essa exclusão, a amostra final de 346 profissionais foi submetida à verificação de valores omissos. Ressalta-se que, embora o cabeçalho do relatório estatístico tenha registrado 378 participantes no arquivo original, a análise efetiva restringiu-se aos 346 casos selecionados como válidos e completos.

A análise fatorial confirmatória (AFC) foi executada nos softwares *Factor* e *JASP*, utilizando a abordagem *Mild-Restricted Confirmatory Factor Analysis* (Ferrando & Lorenzo-Seva, 2017; JASP Team, 2025). Este método é indicado para modelos que buscam maior flexibilidade e robustez na estimação de parâmetros em escalas de autorrelato. Para a correção do qui-quadrado robusto, aplicou-se a técnica LOSEFER, que assegura estimativas mais precisas mesmo em distribuições que se afastam da normalidade multivariada. O uso de matrizes policóricas justifica-se pela natureza ordinal dos itens (escala Likert), permitindo uma estimação mais fidedigna da estrutura latente (Ferrando & Lorenzo-Seva, 2017).

A adequação do modelo confirmatório foi avaliada pelos seguintes índices de ajuste: *Comparative Fit Index* (CFI), *Tucker-Lewis Index* (TLI/NNFI) e *Root Mean Square Error of Approximation* (RMSEA). Seguindo as recomendações da literatura (Brown, 2015; Hair et al., 2014), valores de *CFI* e *TLI* iguais ou superiores a 0,95 indicam um ajuste excelente, enquanto o *RMSEA* deve situar-se abaixo de 0,08 para ser considerado aceitável, sendo o patamar ideal abaixo de 0,06.

Validade e confiabilidade dos construtos

Para as evidências de validade externa e convergente, incluiu-se na coleta a dimensão “crenças” do instrumento de Gonçalves e Leite (2005). As análises confirmatórias adicionais e correlações foram conduzidas nos softwares *Factor* e *JASP* (Ferrando & Lorenzo-Seva, 2017; JASP Team, 2025), assegurando a precisão dos parâmetros estimados e a congruência teórica entre os constructos analisados.

A confiabilidade foi avaliada pelo coeficiente Alfa de Cronbach (α) e pelo coeficiente Ômega de McDonald (ω), sendo este último considerado uma medida mais robusta para escalas com natureza ordinal. A validade do fator é expressa pelo tamanho

das cargas fatoriais, ou seja, quanto maiores elas forem, mais a variável representa o construto (Pasquali, 2013). Além disso, a estabilidade do fator foi verificada pelo índice *H-Latente*, que descreve o grau em que os itens indicam o mesmo construto latente (Ferrando & Lorenzo-Seva, 2018). A validade convergente expressa que os itens de um construto devem compartilhar elevada variância comum (Hair et al., 2014).

As cargas padronizadas precisam estar acima de 0,50 para serem consideradas aceitáveis (Hair et al., 2014), a confiabilidade (α e ω) deve estar acima de 0,70 e a Variância Média Extraída (VME) precisa ser igual ou superior a 0,50 (Fornell & Larcker, 1981; Hair et al., 2014).

Resultados

A análise das entrevistas resultou na identificação de oito categorias centrais que definem a efetividade da AD no contexto investigado. Os resultados indicam que, para os participantes, um sistema efetivo transcende a técnica, exigindo clareza estrutural, utilidade estratégica, retorno motivacional através do *feedback* e suporte tecnológico para a geração de dados confiáveis. Essas dimensões serviram de base para a redação dos itens iniciais do instrumento, garantindo que a escala contemplasse tanto os desafios processuais quanto o impacto organizacional percebido. A análise de conteúdo dos grupos focais revelou oito categorias centrais, que variaram desde a “estrutura e objetividade” até o “impacto estratégico e tomada de decisão”. Nas verbalizações, os participantes destacaram a AD como um “grande diagnóstico” e a necessidade de superar a visão da avaliação como “apenas uma burocracia”. Curiosamente, embora as percepções qualitativas fossem multifacetadas, as análises fatoriais demonstraram que esses elementos convergem para uma estrutura unidimensional.

Tabela 1

Categorias analíticas de efetividade da AD emergentes dos grupos focais.

Categoria	Definição / Achado Principal
Aplicação e Relevância da Avaliação	A aplicação e relevância da avaliação dizem respeito a sua conexão com a realidade e aos objetivos da organização, garantindo que o processo avaliativo seja aplicável e coerente com o contexto organizacional. Envolve a identificação da necessidade de diagnóstico e mensuração prévia do que será avaliado, bem como o mapeamento de competências e a definição de indicadores que assegurem a efetividade e a utilidade da avaliação para o desenvolvimento individual e institucional.
Desafios e Melhorias no Processo Avaliativo	Os desafios e melhorias no processo avaliativo refere-se às limitações e oportunidades de aprimoramento no processo avaliativo, considerando aspectos como a burocracia envolvida, a abrangência e a pertinência do conteúdo avaliado, além da efetiva participação de funcionários e líderes. Também aborda a percepção dos participantes quanto à capacidade do instrumento de avaliação em promover mudanças significativas no futuro.
Desenvolvimento e Gestão de Talentos	O desenvolvimento e gestão de talentos refere-se a gestão de desempenho; ao desenvolvimento de competências e conhecimentos durante a jornada do participante; e às estratégias de planejamento de carreira, que envolvem promoção, retenção e sucessão. Além da construção de relacionamentos entre avaliador e avaliado.
Engajamento, Motivação e Feedback decorrentes da avaliação de desempenho	O engajamento, motivação e feedback refere-se ao processo no qual a avaliação de desempenho é utilizada para engajar e motivar os participantes, oferecer feedbacks claros e construtivos que permitam reconhecer resultados, diagnosticar necessidades de aprimoramento e alinhar o desempenho e competências à cultura organizacional.
Estratégia e Impacto Organizacional	A estratégia e impacto organizacional refere-se à capacidade do processo avaliativo de gerar efeitos concretos e mensuráveis nos níveis estratégico, tático e operacional da organização. Envolve o alinhamento entre a avaliação de desempenho e os objetivos organizacionais, destacando o papel do RH como agente estratégico e integrador. Avaliar as percepções sobre como os resultados das avaliações orientam decisões de gestão de pessoas e afetam a cultura organizacional.
Estrutura e Objetividade na Avaliação	A estrutura e a objetividade na avaliação dizem respeito à clareza dos propósitos do processo avaliativo, à transparência em sua condução e à comunicação eficaz com os participantes. Envolve a definição objetiva dos critérios, o alinhamento entre avaliadores e avaliados quanto às etapas e expectativas da avaliação, garantindo que todos compreendam o processo e seus objetivos de forma clara e acessível.
Percepção e Autoavaliação no Processo Avaliativo	A percepção e a autoavaliação no processo avaliativo dizem respeito à capacidade do indivíduo de refletir criticamente sobre seu próprio desempenho, compreender os critérios utilizados na avaliação e participar ativamente da gestão do próprio desenvolvimento. Envolve tanto a análise subjetiva de suas competências e resultados quanto a percepção dos parâmetros definidos para a mensuração e o diagnóstico do que será avaliado.
Uso de Dados e Tecnologia na Avaliação	O uso de dados e tecnologia na avaliação refere-se à tomada de decisões fundamentadas nas informações obtidas ao longo do processo avaliativo, ao uso de ferramentas tecnológicas que otimizam e facilitam a aplicação das avaliações, bem como à análise do retorno sobre o investimento (ROI) em ações de avaliação e treinamento, a partir dos dados coletados. Essa abordagem fortalece a eficácia do processo avaliativo e orienta estratégias de desenvolvimento mais precisas e embasadas.

A Tabela 1 apresenta as oito categorias analíticas derivadas da análise de conteúdo das transcrições dos grupos focais. Observou-se que a percepção de efetividade dos participantes não se limita à precisão técnica da nota, mas expande-se para a utilidade do *feedback* e o impacto estratégico da ferramenta. Essas categorias foram essenciais, servindo como base para a redação dos itens que compuseram a versão inicial da EEAD.

Na etapa de validação por juízes adotou-se o critério de exclusão para itens que apresentassem CVC inferior a 0,50, seguindo as recomendações de Hernández-Nieto (2002). Após este refinamento, 25 itens foram descartados por não atingirem o índice mínimo ou por apresentarem redundância teórica, resultando em uma versão de 47 itens, mais 4 itens foram retirados posteriormente por estarem duplicados, totalizando 43 itens finais. Os itens retidos apresentaram índices satisfatórios em todos os critérios, e o CVC geral do instrumento alcançou o valor de 0,68.

Tabela 2

Cálculo do CVC dos Itens da Escala.

N.º	CVC (Médias por Critério)			CVC Item
	CL	PP	RT	
1	4,00	4,00	4,00	0,50
2	3,50	4,00	3,75	0,33
3	3,67	4,17	3,92	0,50
4	4,50	4,17	4,33	0,67
5	3,83	4,00	3,92	0,67
6	3,33	4,00	3,67	0,33
7	4,50	4,17	4,33	0,67
8	3,33	3,83	3,58	0,33

9	2,67	3,83	3,25	0,17
10	4,00	5,00	4,50	0,83
11	4,17	5,00	4,58	0,83
12	3,83	5,00	4,42	0,67
13	4,00	4,50	4,25	0,67
14	3,33	4,67	4,00	0,67
15	3,83	4,50	4,17	0,50
16	3,83	4,33	4,08	0,67
17	3,00	5,00	4,00	0,50
18	3,33	5,00	4,17	0,67
19	4,17	4,17	4,17	0,67
20	4,00	4,50	4,25	0,67
21	3,33	3,83	3,58	0,50
22	3,17	3,83	3,50	0,50
23	3,50	3,33	3,42	0,50
24	4,50	4,33	4,42	0,83
25	4,33	4,50	4,42	0,83
26	3,33	3,67	3,50	0,50
27	4,33	3,17	3,75	0,50
28	4,17	4,00	4,08	0,50
29	3,83	4,17	4,00	0,50
30	4,17	4,17	4,17	0,67
31	3,67	4,17	3,92	0,50
32	4,50	4,17	4,33	0,67
33	4,33	4,17	4,25	0,67
34	4,50	4,17	4,33	0,67
35	4,50	4,17	4,33	0,67
36	4,50	4,17	4,33	0,67
37	3,67	2,83	3,25	0,33
38	3,67	3,83	3,75	0,50
39	4,00	3,83	3,92	0,50
40	4,00	3,83	3,92	0,50
41	3,67	3,50	3,58	0,50

42	3,83	3,67	3,75	0,50
43	4,00	3,33	3,67	0,33
44	3,67	3,83	3,75	0,50
45	3,50	3,83	3,67	0,50
46	4,33	5,00	4,67	1,00
47	4,67	5,00	4,83	1,00
48	4,33	5,00	4,67	0,83
49	4,50	5,00	4,75	1,00
50	4,67	5,00	4,83	1,00
51	4,33	4,50	4,42	0,83
52	4,33	5,00	4,67	0,83
53	4,83	5,00	4,92	1,00
54	4,50	5,00	4,75	0,83
55	5,00	5,00	5,00	1,00
56	4,33	5,00	4,67	0,83
57	4,83	5,00	4,92	1,00
58	5,00	5,00	5,00	1,00
59	5,00	5,00	5,00	1,00
60	5,00	5,00	5,00	1,00
61	4,83	4,67	4,75	0,83
62	4,17	5,00	4,58	0,83
63	4,83	5,00	4,92	1,00
64	4,33	4,33	4,33	0,83
65	3,83	4,33	4,08	0,67
66	3,83	4,33	4,08	0,67
67	3,83	4,33	4,08	0,67
68	3,83	4,33	4,08	0,67
69	4,33	4,33	4,33	0,83
70	3,83	4,33	4,08	0,67
71	4,50	4,33	4,42	0,83
72	4,50	4,33	4,42	0,83
Média Geral	4,08	4,34	4,21	0,68

Nota. N.º = número do item; CVC = coeficiente de validade de conteúdo; CL = clareza de linguagem; PP = pertinência prática; RT = relevância teórica; CVC Item = coeficiente de validade de conteúdo por item. As médias foram calculadas com base na avaliação de juízes.

A amostra final da AFE foi composta por 483 profissionais, atendendo ao critério de tamanho amostral recomendado por Pasquali (2013), que sugere um mínimo de 10 respondentes por item do instrumento. O perfil dos participantes é formado, em sua maioria, por mulheres (71,2%), atuantes em organizações de natureza privada (67,9%) e com idade situada na faixa entre 26 e 35 anos (35,8%). No que diz respeito à formação acadêmica, a maioria possui ensino superior completo (49,3%) ou pós-graduação (35,6%). Quanto ao tempo de serviço na organização atual, destaca-se o grupo com permanência entre 1 e 3 anos (26,1%), conforme detalhado na Tabela 3.

Tabela 3

Caracterização da amostra.

Variável / Categoria	n	%
Tipo de Organização		
Privada	328	67,9%
Pública	75	15,5%
Outro	38	7,9%
Mista	25	5,2%
Terceiro Setor / Sem Fins Lucrativos	17	3,5%
Gênero		
Feminino	344	71,2%
Masculino	136	28,2%
Outro	2	0,4%
Prefiro não responder	1	0,2%
Idade		

Até 25 anos	99	20,5%
26 a 35 anos	173	35,8%
36 a 45 anos	139	28,8%
46 a 55 anos	62	12,8%
56 anos ou mais	10	2,1%
Escolaridade		
Ensino fundamental	2	0,4%
Ensino médio	71	14,7%
Ensino superior	238	49,3%
Pós-graduação	172	35,6%
Tempo na Organização		
Menos de 1 ano	96	19,9%
De 1 a 3 anos	126	26,1%
De 3 a 5 anos	110	22,8%
De 5 a 10 anos	78	16,1%
Mais de 10 anos	73	15,1%

Para verificar a adequação da matriz de correlações policóricas para a análise fatorial, foram calculados os índices de Kaiser-Meyer-Olkin (KMO) e o Teste de Esfericidade de Bartlett. O KMO apresentou valor de 0,984, considerado excelente, e o Teste de Bartlett foi estatisticamente significativo ($\chi^2(903) = 19137,9; p < 0,001$), indicando que os dados eram adequados para a fatoração.

Após a indicação da Análise Paralela para a retenção de uma única dimensão, procedeu-se à avaliação dos índices de ajuste do modelo unifatorial, uma vez que o programa Factor disponibiliza estatísticas de ajuste que permitem uma análise rigorosa da estrutura proposta (Ferrando & Lorenzo-Seva, 2017).

Tabela 4

Variância explicada com base nos autovalores e análise paralela.

Variável	Autovalor	Proporção de Variância	Proporção Cumulativa de Variância	Percentual de variância dos dados reais Análise Paralela
----------	-----------	------------------------	-----------------------------------	--

1	27,04	0,63	0,63	65,10
2	1,76	0,04	0,67	4,20
3	1,46	0,03	0,70	3,43

Nota. O número de fatores a ser retido é um, pois apenas um fator dos dados reais apresenta a porcentagem de variância explicada maior do que os dados aleatórios.

Conforme apresenta a Tabela 5, o modelo com um fator demonstrou indicadores de ajuste excelentes, apresentando uma baixa proporção entre o valor de χ^2 (qui-quadrado) e os graus de liberdade (gl), além de um *RMSEA* inferior a 0,05 e valores de *CFI* e *TLI* superiores a 0,99 (Hair et al., 2014; Kline, 2016).

Tabela 5

Estatísticas de Ajuste para o Modelo Unifatorial.

	χ^2	df	χ^2/df	RMSEA	CFI	TLI
1 fator	1766,60	860	2,05	0,047	0,997	0,997

A Tabela 6 apresenta a extração do fator único com as respectivas cargas fatoriais dos itens que compõem o modelo final. Adotou-se o critério de corte de 0,778 para as cargas fatoriais, visando assegurar que apenas os itens com forte contribuição para a variável latente fossem mantidos (Hair et al., 2014). Do total inicial de 43 itens submetidos à análise, 12 itens (V7, V9, V10, V11, V12, V15, V28, V36, V38, V39, V40 e V42) foram excluídos por apresentarem cargas inferiores ao limite estabelecido.

Os 31 itens retidos apresentaram cargas fatoriais elevadas, variando entre 0,779 e 0,878. Como a solução sugerida pela Análise Paralela foi unifatorial, não foram observados problemas de cargas cruzadas entre dimensões, garantindo a unidimensionalidade e a clareza teórica do instrumento para a etapa confirmatória.

A rotação utilizada foi a *Robust Promin* (Lorenzo-Seva & Ferrando, 2019). Adotou-se um critério de corte de 0,778 para as cargas fatoriais, resultando na exclusão de itens com carga inferior e mantendo 31 itens para a etapa confirmatória. A adequação

do modelo na fase exploratória apresentou índices excelentes: $RMSEA = 0,047$; $CFI = 0,997$ e $TLI = 0,997$. A estabilidade do fator foi avaliada por meio do índice H (Hancock & Mueller, 2000), com valor de H -latente de 0,988, sugerindo uma variável latente bem definida e estável. A confiabilidade do instrumento foi atestada pelo índice Ômega (ω) (McDonald, 1999), que apresentou valor de 0,986.

Tabela 6

Cargas Fatoriais da Matriz de Fator Único.

Itens	Carga Fatorial (λ)	Comunalidade (h^2)	Singularidade (u^2)
Item 1	0,779	0,607	0,393
Item 2	0,800	0,640	0,360
Item 3	0,808	0,653	0,347
Item 4	0,814	0,662	0,338
Item 5	0,819	0,672	0,328
Item 6	0,804	0,647	0,353
Item 8	0,828	0,686	0,314
Item 13	0,862	0,742	0,258
Item 14	0,874	0,764	0,236
Item 16	0,841	0,708	0,292
Item 17	0,839	0,704	0,296
Item 18	0,834	0,696	0,304
Item 19	0,869	0,755	0,245
Item 20	0,801	0,641	0,359
Item 21	0,837	0,701	0,299
Item 22	0,872	0,761	0,239
Item 23	0,842	0,709	0,291
Item 24	0,842	0,708	0,292
Item 25	0,833	0,695	0,305
Item 26	0,837	0,701	0,299

Item 27	0,780	0,609	0,391
Item 29	0,827	0,684	0,316
Item 30	0,827	0,683	0,317
Item 31	0,878	0,771	0,229
Item 32	0,807	0,652	0,348
Item 33	0,825	0,680	0,320
Item 34	0,821	0,674	0,326
Item 35	0,808	0,654	0,346
Item 37	0,779	0,607	0,393
Item 41	0,779	0,608	0,392
Item 43	0,798	0,638	0,362

Os índices de replicabilidade dos escores fatoriais *H-latente* e *H-observed* foram, respectivamente, 0,988 e 0,986. Ambos os valores situam-se significativamente acima de 0,80, indicando que a variável latente está bem definida e possui elevada estabilidade, o que sugere que resultados semelhantes serão obtidos em estudos futuros (Ferrando & Lorenzo-Seva, 2018; Hancock & Mueller, 2001). As comunalidades (h^2) dos itens variaram entre 0,607 e 0,771, evidenciando que o fator latente explica uma parcela substancial da variância de cada indicador (superior a 60% em todos os casos). Em contrapartida, as singularidades (u^2) mantiveram-se em patamares reduzidos, oscilando entre 0,229 e 0,393. Esses resultados ratificam a precisão dos indicadores e a baixa incidência de erro de medida, confirmando a robustez da estrutura interna da Escala de Efetividade da Avaliação de Desempenho (EEAD).

Os itens da escala foram integralmente respondidos por 378 pessoas, sendo que a amostra final utilizada para as análises confirmatórias foi composta por 346 participantes após o tratamento de dados, atendendo ao critério de tamanho amostral recomendado por Pasquali (2013), que sugere um mínimo de 10 respondentes por item

do instrumento. A amostra, como apresentado na Tabela 7, é formada em sua maioria por mulheres, atuantes em organizações privadas, com ensino superior completo, na faixa etária entre 26 e 35 anos e com tempo de serviço na organização atual entre 5 e 10 anos. Houve uma predominância acentuada de respondentes do setor privado (75,1%) e do gênero feminino (75,1%).

Tabela 7

Caracterização da amostra.

Variável / Categoria	n	%
Tipo de Organização		
Privada	284	75,1%
Pública	44	11,6%
Outro	38	10,1%
Mista	5	1,3%
Terceiro Setor / Sem Fins Lucrativos	7	1,9%
Gênero		
Feminino	284	75,1%
Masculino	89	23,5%
Outro	4	1,1%
Prefiro não responder	1	0,3%
Idade		
Até 25 anos	55	14,6%
26 a 35 anos	172	45,5%
36 a 45 anos	113	29,9%
46 a 55 anos	30	7,9%
56 anos ou mais	8	2,1%
Escolaridade		
Ensino fundamental	4	1,1%
Ensino médio	52	13,8%
Ensino superior	189	50,0%
Pós-graduação	133	35,2%

Tempo na Organização		
Menos de 1 ano	43	11,4%
De 1 a 3 anos	68	18,0%
De 3 a 5 anos	110	29,1%
De 5 a 10 anos	113	29,9%
Mais de 10 anos	44	11,6%

Constatou-se a ausência total de valores omissos (0%) nesta amostra final, conforme atestado pelo bloco de descritivos de dados perdidos (*Missing Data*). Dessa forma, apesar de o método *Hot-Deck Multiple Imputation* (Lorenzo-Seva & Van Ginkel, 2016) estar parametrizado no protocolo de análise para garantir a robustez da matriz, não houve necessidade de processamento de imputação, visto que a base de dados analisada encontrava-se integralmente preenchida. A adequação da matriz de correlações policóricas também foi verificada. O índice de *Kaiser-Meyer-Olkin* (KMO) apresentou valor de 0,972, e o Teste de Esfericidade de Bartlett revelou-se significativo ($\chi^2 [465] = 3843,4; p < 0,001$), confirmando a fatorabilidade da matriz.

Foi testado o modelo unidimensional para a Escala de Efetividade da Avaliação de Desempenho (EEAD), composto por 31 itens, conforme derivado da etapa exploratória anterior. O modelo apresentou índices de ajuste excelentes ($CFI = 0,993$; $TLI = 0,992$ e $RMSEA = 0,064$; $IC\ 90\% 0,060 - 0,069$), confirmando a estrutura teórica proposta. Adicionalmente, foram calculados os índices de fidedignidade por meio do Alfa de Cronbach e do Ômega de McDonald (ω), bem como a estabilidade do fator pelo índice *H-Latente*.

Tabela 8

Índices de ajuste do modelo testado.

Modelos	χ^2	df	χ^2/df	RMSEA	IC 90% RMSEA	CFI	TLI
M1 - 1 fator	1051,94	434	2,42	0,064	[0,060 – 0,069]	0,993	0,992

Assim, a AFC confirmou o instrumento com uma estrutura unidimensional. A escala foi validada em sua configuração de 31 itens, apresentando excelentes índices de ajuste.

Em complemento à avaliação dos índices de ajuste global, procedeu-se à análise da validade local do modelo de mensuração e da precisão dos indicadores (Tabela 9). As cargas fatoriais padronizadas (λ) apresentaram magnitudes elevadas, situando-se entre 0,713 e 0,943, o que indica uma forte associação entre os itens e o construto de efetividade da AD. No que tange às comunalidades (h^2), os valores oscilaram entre 0,542 e 0,890, evidenciando que o fator latente explica mais de 54% da variância de cada item individualmente, patamar este considerado excelente na literatura. Em contrapartida, as singularidades (u^2), que representam a variância residual ou o erro de medida, mantiveram-se em níveis reduzidos, variando entre 0,110 e 0,458. A robustez da estrutura foi ratificada pela Variância Média Extraída (VME) de 0,65, valor que supera o ponto de corte de 0,50 (Hair et al., 2014) e confirma a validade convergente do instrumento. Por fim, a fidedignidade da escala foi atestada pelos elevados índices de consistência interna e replicabilidade, incluindo o Ômega de McDonald ($\omega = 0,986$), Alfa de Cronbach ($\alpha = 0,98$) e o H-Latente ($H = 0,988$).

Tabela 9

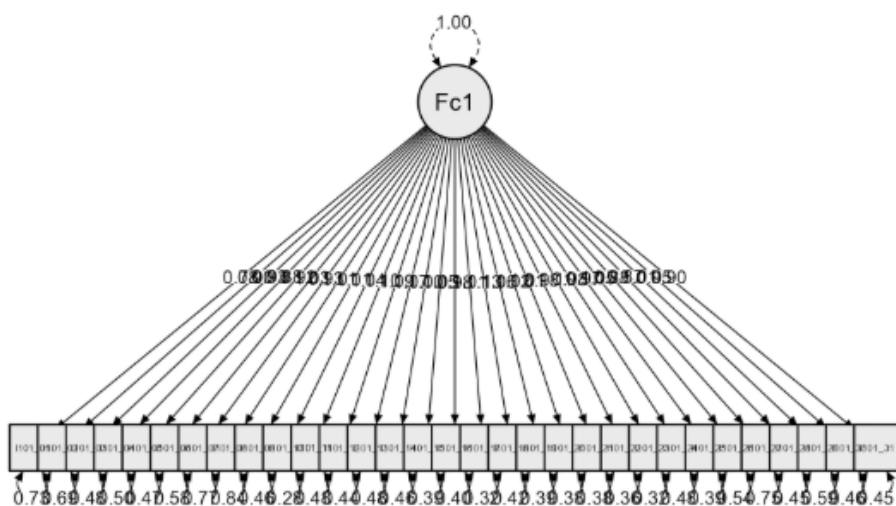
Cargas fatoriais padronizadas.

Itens	Carga Fatorial (λ)	Comunalidade (h^2)	Singularidade (u^2)
Item 1	0,746	0,557	0,443
Item 2	0,802	0,644	0,356
Item 3	0,822	0,676	0,324
Item 4	0,793	0,630	0,370
Item 5	0,799	0,638	0,362
Item 6	0,848	0,719	0,281
Item 7	0,736	0,542	0,458

Item 8	0,797	0,636	0,364
Item 9	0,843	0,710	0,290
Item 10	0,943	0,890	0,110
Item 11	0,892	0,796	0,204
Item 12	0,896	0,802	0,198
Item 13	0,894	0,799	0,201
Item 14	0,896	0,803	0,197
Item 15	0,897	0,833	0,167
Item 16	0,896	0,799	0,201
Item 17	0,898	0,757	0,243
Item 18	0,887	0,725	0,275
Item 19	0,923	0,840	0,160
Item 20	0,898	0,824	0,176
Item 21	0,864	0,827	0,173
Item 22	0,887	0,809	0,191
Item 23	0,893	0,880	0,120
Item 24	0,835	0,697	0,303
Item 25	0,795	0,750	0,250
Item 26	0,824	0,748	0,252
Item 27	0,713	0,630	0,370
Item 28	0,769	0,700	0,300
Item 29	0,799	0,731	0,269
Item 30	0,857	0,801	0,199
Item 31	0,771	0,751	0,249
VME	0,65		
ω	0,986		
α	0,98		
H-Latente	0,988		

Figura 1

Modelo estrutural/path diagram da AFC.

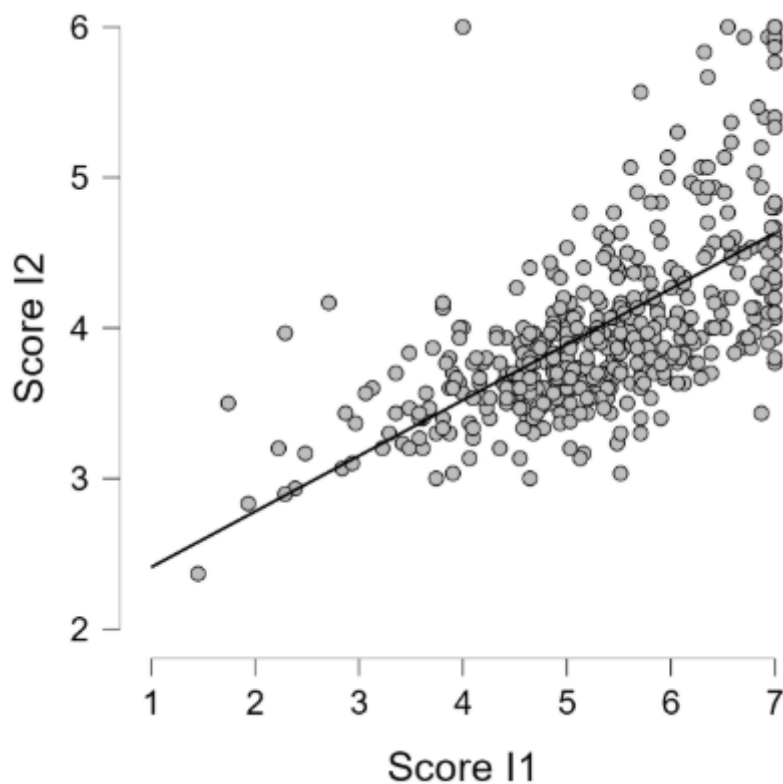


Para verificar a validade convergente do instrumento desenvolvido (Score I1), realizou-se uma análise de correlação com a dimensão “crenças” da escala de Gonçalves e Leite (2005) (Score I2). A escolha da escala de Gonçalves e Leite (2005) justifica-se por ser um instrumento nacional com robustos índices de confiabilidade ($\alpha = 0,90$), desenvolvido para mensurar atitudes frente ao processo de AD em três categorias: crenças, sentimentos e capacitação. Para este estudo, utilizou-se a dimensão “crenças”, que mede a credibilidade e o grau de confiança que o indivíduo deposita no sistema. Segundo as autoras, a crença refere-se ao quanto o trabalhador tem como “certo” ou “verdadeiro” os objetivos da avaliação. Assim, se o profissional acredita na utilidade e imparcialidade do processo, sua atitude tende a ser favorável. A convergência teórica reside no fato de que um sistema de AD só pode ser considerado efetivo se for percebido como legítimo e confiável pelos avaliados, tornando a dimensão de “crenças” um critério externo ideal para validar a nova escala. Inicialmente, o teste de Shapiro-Wilk indicou que as variáveis não apresentavam distribuição normal ($p < ,001$), justificando a priorização do coeficiente de correlação de Spearman.

Os resultados revelaram uma correlação positiva, forte e estatisticamente significativa entre os escores dos instrumentos ($p = 0,67; p < ,001$). O coeficiente de Pearson também apresentou magnitude similar ($p = 0,64; p < ,001$). Esses achados fornecem evidências de validade baseada na relação com variáveis externas, sugerindo que o novo instrumento converge teoricamente com o constructo de crenças organizacionais previamente estabelecido na literatura.

Figura 2

Gráfico de dispersão (Scatter plot) entre os escores do instrumento atual e a dimensão de crenças, apresentando a linha de tendência linear.



O instrumento obteve índices de precisão excepcionais ($\alpha = 0,98$; $\omega = 0,986$) e uma VME de 0,65, atendendo amplamente aos preceitos de adequabilidade prescritos pela literatura. A estabilidade do fator também foi confirmada pelo alto valor do índice H (0,988).

O instrumento final foi composto pelos 31 itens, usando uma escala de concordância de 1 a 7, onde 1 significa “discordo totalmente” e 7 significa “concordo totalmente”.

Tabela 10

Itens da Escala de Efetividade da Avaliação de Desempenho (EEAD).

N.º	Descrição do Item
1	Os objetivos do processo de avaliação de desempenho são claros para mim.

- 2 As diferentes etapas do processo de avaliação são explicadas de forma transparente.
- 3 Os critérios utilizados na minha avaliação são objetivos.
- 4 Os critérios utilizados na minha avaliação são fáceis de entender.
- 5 Compreendo como meu desempenho é mensurado no processo avaliativo.
- 6 Entendo claramente como os resultados da minha avaliação serão utilizados.
- 7 Há um alinhamento claro entre o que é comunicado sobre a avaliação e o que é efetivamente praticado.
- 8 O modelo atual de avaliação de desempenho da organização possibilita oportunidades de melhoria.
- 9 O processo de avaliação de desempenho, como é hoje, é útil para o meu desenvolvimento.
- 10 A avaliação aborda os aspectos que são realmente importantes para a realização do meu trabalho.
- 11 Os critérios utilizados na minha avaliação são coerentes com as atividades do meu cargo.
- 12 A avaliação de desempenho evidencia as competências essenciais para o meu cargo.
- 13 O processo de avaliação ajuda a identificar pontos de melhoria no meu desempenho.
- 14 O processo de avaliação de desempenho está alinhado aos valores da nossa organização.
- 15 A avaliação reforça a importância das competências que a organização espera de mim.
- 16 O processo de avaliação de desempenho estimula a melhoria contínua do meu desempenho.
- 17 O processo de avaliação me ajuda a identificar minhas necessidades de capacitação e desenvolvimento.
- 18 A partir dos resultados da minha avaliação, são propostas ações concretas para o meu desenvolvimento profissional.
- 19 A avaliação de desempenho é uma ferramenta útil para o planejamento da minha carreira na organização.
- 20 O processo de avaliação me ajuda a reconhecer minhas oportunidades de crescimento.
- 21 O processo de autoavaliação é um estímulo para a melhoria do meu desempenho.
- 22 O *feedback* que recebo no processo de avaliação é construtivo.
- 23 O *feedback* que recebo é uma ferramenta útil para o meu desenvolvimento profissional.
- 24 O processo de avaliação gera *feedbacks* claros sobre os comportamentos esperados no meu trabalho.
- 25 Minhas conquistas e resultados são reconhecidos durante o processo de avaliação de desempenho.
- 26 O processo de avaliação de desempenho me motiva a buscar melhores resultados no meu trabalho.
- 27 O processo de avaliação de desempenho me incentiva a refletir sobre meus pontos fortes.
- 28 A ferramenta (formulário/sistema/plataforma) utilizada facilita a aplicação da minha avaliação de desempenho.

- 29 A ferramenta (formulário/sistema/plataforma) utilizada permite acompanhar os resultados da avaliação com facilidade.
 - 30 As informações da minha avaliação são usadas para subsidiar decisões sobre meu desenvolvimento.
 - 31 As informações da avaliação orientam ajustes nas práticas de gestão de pessoas.
-

Discussão

O objetivo deste estudo foi desenvolver e encontrar índices de validade para uma escala que pretende mensurar a efetividade da Avaliação de Desempenho (AD) no contexto organizacional. A integração dos achados qualitativos, provenientes da análise de conteúdo dos grupos focais, com os resultados psicométricos, permitiu a construção de um instrumento que reflete a complexidade das percepções dos profissionais sobre o que torna uma avaliação verdadeiramente efetiva.

Embora o roteiro dos grupos focais tenha explorado dimensões previamente mapeadas na literatura, a análise qualitativa revelou que, na percepção dos trabalhadores, tais elementos não operam de forma isolada. Observou-se uma convergência espontânea nos relatos, nos quais a utilidade do *feedback*, o suporte tecnológico e a transparência dos critérios foram articulados como componentes interdependentes de um mesmo fenômeno. Esse caráter integrado da efetividade da AD, manifestado no discurso dos participantes, encontrou sustentação estatística robusta na etapa quantitativa, na qual a estrutura unidimensional do instrumento explicou 65,1% da variância total. Nesse sentido, é importante destacar que, embora o roteiro tenha servido como disparador das discussões, a força das cargas fatoriais e a variância explicada confirmaram que esses temas pertencem, de fato, a um único núcleo de sentido para o trabalhador. Portanto, a unidimensionalidade não decorre apenas de uma escolha metodológica, mas reflete a essência da “utilidade percebida” pelo sujeito. Para o

avaliado, a ferramenta só é considerada efetiva quando todos os seus componentes operam em harmonia, validando o construto como uma unidade psicológica coesa.

Um tema recorrente na fase qualitativa foi o engajamento, motivação e *feedback*, apontando que “rodar uma avaliação por si só não é efetivo”. A escala reflete essas falas ao apresentar altos índices de fidedignidade ($\omega = 0,98$), indicando que ela é capaz de medir de forma precisa se esses rituais pós-avaliação estão ocorrendo.

Os resultados deste estudo corroboram as evidências clássicas de que a efetividade da AD está mais ligada à qualidade do processo do que à sofisticação técnica do sistema. Assim como proposto por Findley et al. (2000), observou-se que o *feedback* e o planejamento da sessão são preditores fundamentais da percepção de valor da ferramenta. A fala recorrente nos grupos focais de que “rodar uma avaliação por si só não é efetivo” alinha-se aos achados de Longenecker et al. (1988), reforçando que a efetividade depende da utilidade prática percebida tanto pelo gestor quanto pelo liderado.

A correlação forte e positiva com a dimensão de “crenças” de Gonçalves e Leite (2005) ($p = 0,67$; $p < 0,001$) ratifica a perspectiva de Sharma et al. (2016) e Modipane et al. (2019), que apontam a justiça e a transparência como pilares da efetividade. Quando o trabalhador acredita na imparcialidade do processo, a AD deixa de ser um rito burocrático e passa a ser vista como um subsídio para o desenvolvimento, o que corrobora a tese de Iqbal et al. (2019) sobre o critério de utilização: sistemas voltados para o desenvolvimento geram maior percepção de efetividade do que aqueles focados meramente em punição ou controle administrativo.

Por outro lado, os dados apresentam uma nuance distinta de modelos multidimensionais como o de Ikramullah et al. (2016). Enquanto esses autores sugerem

que a efetividade deve ser medida por quatro quadrantes distintos (processos internos, relações humanas, sistema aberto e objetivos racionais), a estrutura da EEAD revelou-se unidimensional. Isso sugere que, no contexto laboral brasileiro investigado, o trabalhador não fragmenta a experiência da AD, ele a percebe de forma integrada. Se existe falha no *feedback* ou na tecnologia, o sentimento de ineficácia contamina o construto como um todo.

Os dados demográficos revelam uma amostra qualificada (84,9% com ensino superior ou pós-graduação) e majoritariamente do setor privado (67,9%). Para este perfil de trabalhador, a efetividade da AD está ligada à capacidade da ferramenta em gerar planos de ação e desenvolvimento, e não apenas em preencher formulários. As verbalizações sobre “ROI (retorno sobre investimento) e treinamentos” e “tomada de decisão” indicam que a AD deve servir como um subsídio para o RH Estratégico.

A escala permite que as empresas saiam do campo da subjetividade para o campo da evidência científica. Com índices de ajuste excelentes ($CFI = 0,993$; $TLI = 0,992$), o instrumento oferece segurança psicométrica para ser utilizado como ferramenta de auditoria de processos de RH, permitindo identificar se o investimento de tempo e recursos na avaliação está gerando a percepção de valor esperada.

Assim, a fidedignidade elevada ($\omega = 0,98$) e a variância explicada sugerem que a EEAD captura a essência dessa “utilidade percebida”. Para uma amostra qualificada (84,9% com ensino superior), a efetividade não é um conceito abstrato, mas um resultado tangível traduzido em planos de ação e suporte ao RH (Recursos Humanos) estratégico. O instrumento, portanto, oferece segurança psicométrica ($CFI = 0,993$) para que as organizações realizem auditorias baseadas em evidências científicas, superando a lacuna entre a intenção do sistema e a realidade percebida.

Conclusão

O presente estudo cumpriu seu objetivo principal ao desenvolver e encontrar indícios de validade da Escala de Efetividade da Avaliação de Desempenho (EEAD). A pesquisa demonstrou que a efetividade, embora dependa de múltiplos processos técnicos, é percebida como um construto unitário pelo trabalhador, consolidando-se como uma medida robusta e confiável para o contexto organizacional brasileiro.

Contribuições do Estudo

As contribuições deste trabalho são tanto teóricas quanto práticas. No campo acadêmico, o estudo inova ao identificar a unidimensionalidade da efetividade percebida. Enquanto modelos internacionais, como o de Ikramullah et al. (2016), propõem divisões multifatoriais, esta pesquisa revela que, na percepção do empregado, os elementos da AD (*feedback*, tecnologia, critérios) operam de forma integrada. No campo prático, a EEAD oferece aos gestores uma ferramenta de auditoria de baixo custo e alta precisão psicométrica. Ela permite que o RH saia do campo da intuição para o campo da evidência científica, identificando se o investimento de tempo e recursos nos ciclos de avaliação está, de fato, gerando percepção de valor e utilidade estratégica.

Corroboração com a Literatura

Os achados deste estudo corroboram a literatura clássica e contemporânea sobre gestão de desempenho. A importância central do *feedback* e do planejamento das sessões alinha-se aos pressupostos de Findley et al. (2000) e Longenecker (1988). Além disso, a forte convergência encontrada com o sistema de crenças organizacionais ratifica as teses de Sharma et al. (2016) e Modipane et al. (2019), reforçando que a efetividade de qualquer política de RH está condicionada à sua legitimidade e à justiça percebida

pelo avaliado. O estudo confirma, ainda, a visão de Iqbal et al. (2019) de que a AD atinge sua efetividade máxima quando é percebida como um subsídio para o desenvolvimento e não apenas como um rito burocrático de controle.

Limitações do Estudo

Contudo, o estudo apresenta limitações que devem ser consideradas. A amostra foi composta predominantemente por profissionais com alto nível de escolaridade (84,9% com ensino superior ou pós-graduação) e atuantes no setor privado (67,9%). Esse perfil pode não refletir as nuances de contextos laborais menos estruturados tecnologicamente ou com culturas organizacionais distintas, como o setor público ou o terceiro setor. Além disso, o caráter transversal da coleta de dados impede a observação da estabilidade da percepção de efetividade ao longo de múltiplos ciclos de avaliação.

Agenda de Pesquisa

Para futuras investigações, sugere-se a aplicação da EEAD em amostras de nível operacional e em diferentes setores da economia para testar a invariância da medida, garantindo que o instrumento mantenha sua estabilidade independentemente do nível hierárquico. Recomenda-se, também, a exploração da validade preditiva da escala, correlacionando os escores de efetividade percebida com indicadores objetivos de desempenho, tais como cumprimento de metas, taxas de *turnover* e produtividade individual. Tais avanços consolidariam a EEAD não apenas como um diagnóstico de percepção, mas como um ativo de inteligência de negócios capaz de prever resultados organizacionais tangíveis.

Referências

- Bardin, L. (2016). *Análise de conteúdo* (L. A. Reto & A. Pinheiro, Trads.). Edições 70.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.
- Cassepp-Borges, V., Balbinotti, M. A. A., & Teodoro, M. L. M. (2010). Tradução e validação de conteúdo: Uma proposta para a adaptação de instrumentos. In L. Pasquali (Org.), *Instrumentação psicológica: fundamentos e práticas* (pp. 506–520). Artmed.
- Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement, 78*(5), 762–780.
<https://doi.org/10.1177/0013164417719308>
- Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program Factor at 10: Origins, development and future directions. *Psicothema, 29*(2), 236–240.
- Findley, H. M., Giles, W. F., & Mossholder, K. W. (2000). Performance appraisal process and system facets: Relationships with contextual performance. *Journal of Applied Psychology, 85*(4), 634–640.
<https://doi.org/10.1037/0021-9010.85.4.634>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39. <https://doi.org/10.2307/3151312>
- Girdharwal, N. (2019). A comparative study of performance appraisal on effectiveness and organizational commitment. *Prabandhan: Indian Journal of Management, 12*(11), 47–54. <https://doi.org/10.17010/pijom/2019/v12i11/148409>

- Gonçalves, V. L. M., & Leite, M. M. J.. (2005). Instrumento para mensuração de atitudes frente ao processo de avaliação de desempenho. *Revista Brasileira De Enfermagem*, 58(5), 563–567.
<https://doi.org/10.1590/S0034-71672005000500012>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson Education.
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future—A Festschrift in honor of Karl Jöreskog* (pp. 195–216). Scientific Software International.
- Hernandez-Nieto, R. A. (2002). *Contributions to statistical analysis: The coefficients of proportional variance, content validity and kappa*. CreateSpace.
- Ikramullah, M., Van Prooijen, J., Iqbal, M. Z., & Ul-Hassan, F. S. (2016). Effectiveness of performance appraisal: Developing a conceptual framework using competing values approach. *Personnel Review*, 45(2), 334–352.
<https://doi.org/10.1108/PR-07-2014-0164>
- Iqbal, M. Z., Akbar, S., Budhwar, P., & Shah, S. Z. A. (2019). Effectiveness of performance appraisal: Evidence on the utilization criteria. *Journal of Business Research*, 101, 285–299. <https://doi.org/10.1016/j.jbusres.2019.04.035>
- JASP Team. (2025). *JASP* (Versão 0.19.2) [Software de computador].
<https://jasp-stats.org/>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Longenecker, C. O., Liverpool, P. R., & Wilson, K. Y. (1988). An assessment of manager/subordinate perceptions of performance appraisal effectiveness. *Journal*

of Business and Psychology, 2(4), 311–320.

<https://doi.org/10.1007/BF01013763>

- Longo, F. (2007). *Mérito e flexibilidade: A gestão das pessoas no setor público* (A. Corbisier, L. Jahn, L. R. Gil, & P. A. Barbosa, Trads.). Edições Fundap. (Obra original publicada em 2004)
- Lorenzo-Seva, U., & Ferrando, P. J. (2019). *Robust Promin: A method for diagonally weighted factor rotation* [Technical report]. Universitat Rovira i Virgili.
- Lorenzo-Seva, U., & Van Ginkel, J. R. (2016). Multiple imputation of missing values in exploratory factor analysis of multidimensional scales: Estimating latent trait scores. *Anales de Psicología / Annals of Psychology*, 32(2), 596–608.
<https://doi.org/10.6018/analesps.32.2.215161>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
- Modipane, P., Botha, P., & Blom, T. (2019). Employees' perceived effectiveness of the performance management system at a North-West provincial government department. *SA Journal of Human Resource Management*, 17, 12 pages.
<https://doi.org/10.4102/sajhrm.v17i0.1081>
- Pasquali, L. (2010). *Instrumentação psicológica: fundamentos e práticas*. Artmed.
- Pasquali, L. (2013). *Psicometria: Teoria dos testes na psicologia e na educação* (5ª ed.). Vozes.
- Sharma, N. P., Sharma, T., & Agarwal, M. N. (2016). Measuring employee perception of performance management system effectiveness: Conceptualization and scale development. *Employee Relations*, 38(2), 224–247.
<https://doi.org/10.1108/ER01-2015-0006>

Sharma, N. P., Sharma, T., & Agarwal, M. N. (2022). Relationship between perceived performance management system (PMS) effectiveness, work engagement and turnover intention: Mediation by psychological contract fulfillment.

Benchmarking: An International Journal, 29(9), 2985–3007.

<https://doi.org/10.1108/BIJ-01-2021-0008>

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2),

209–220. <https://doi.org/10.1037/a0023353>

Conclusão

A presente dissertação teve como objetivo central investigar a efetividade da Avaliação de Desempenho (AD) sob uma perspectiva multidimensional, integrando a revisão do estado atual da AD, a síntese dos determinantes de efetividade e o desenvolvimento de um instrumento de medida. O percurso metodológico e teórico realizado nos três manuscritos permitiu uma compreensão profunda de como a AD evoluiu de uma ferramenta puramente técnica para um processo psicossocial complexo.

As evidências consolidadas no primeiro estudo revelaram que a produção científica global na última década (2013-2025) aponta para uma transição paradigmática. Identificou-se que a literatura tem priorizado o processo interacional em detrimento do *design* estrutural dos sistemas, destacando variáveis como a justiça organizacional e a liderança transformacional como pilares para a sustentação de sistemas de AD efetivos.

O segundo estudo aprofundou essa discussão ao identificar os determinantes da efetividade. Concluiu-se que a percepção de justiça interacional e a qualidade do *feedback* são os preditores mais consistentes da aceitação do sistema pelos funcionários. A efetividade percebida emergiu não apenas como um resultado, mas como uma variável mediadora fundamental: sistemas de AD que são percebidos como justos e úteis convertem insumos de gestão em resultados organizacionais tangíveis, como o aumento do comprometimento afetivo e da produtividade.

Para suprir a lacuna instrumental identificada, o terceiro estudo focou no desenvolvimento da Escala de Efetividade da Avaliação de Desempenho (EEAD). Os excelentes índices de ajuste obtidos confirmam que o instrumento é uma ferramenta

robusta e adequada para mensurar o construto. A EEAD permite que gestores identifiquem com precisão em que o sistema de AD necessita de aprimoramento.

Em termos teóricos, esta dissertação contribui para a literatura ao consolidar o conceito de efetividade como um fenômeno percebido e socialmente construído. Praticamente, o trabalho oferece às organizações brasileiras um instrumento que pode orientar intervenções estratégicas para reduzir a percepção de burocracia e aumentar o valor percebido da AD.

Apesar das contribuições, reconhece-se como limitação o recorte transversal das amostras quantitativas, o que impede a inferência de causalidade definitiva. Sugere-se que pesquisas futuras utilizem delineamentos longitudinais para observar como a efetividade percebida se altera após ciclos de *feedback* e intervenções de treinamento de liderança.

Em conclusão, a efetividade da avaliação de desempenho não reside na perfeição técnica do formulário, mas na qualidade das relações e na percepção de justiça que o processo promove. Espera-se que este trabalho sirva de base para que acadêmicos e profissionais da área possam transformar a AD de um rito burocrático em uma prática de gestão genuinamente voltada ao desenvolvimento humano e organizacional.