



Universidade de Brasília – UnB
Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas - FACE
Programa de Pós-Graduação em Administração – PPGA

**Construção e Validação de um Modelo Teórico-Integrativo de Avaliação de Efetividade
da Gestão de Desempenho**

Mariana Lopes de Araújo

Orientador: Dr. Pedro Paulo Murce Menezes Cavalcante

Brasília, DF
2024

**Construção e Validação de um Modelo Teórico-Integrativo de Avaliação de Efetividade
da Gestão de Desempenho**

Mariana Lopes de Araújo

Orientador: Prof. Dr. Pedro Paulo Murce Menezes Cavalcante

Tese de doutorado apresentada ao Programa de Pós-Graduação em Administração da Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas da Universidade de Brasília como requisito para obtenção do título de Doutor em Administração.

Comissão Examinadora:

Prof. Dr. Pedro Paulo Murce Menezes Cavalcante

Orientador (Membro Presidente) (Universidade de Brasília – PPGA/UnB)

Prof. Dr. André Luiz Marques Serrano

Membro Examinador (Universidade de Brasília - UnB)

Prof. Dra. Diva Ester Okazaki Rowe

Membro Externo (Universidade de Federal da Bahia – UFBA)

Prof. Dra. Thaís Zerbini

Membro Externo (Universidade de São Paulo - USP)

Prof. Dra. Gisela Demo

Membro Suplente (Universidade de Brasília - UnB)

Brasília, DF, 18 de dezembro de 2024

Dedicatória

Dedico à Ana Carolina, minha companheira de vida, que está comigo muito antes deste sonho e, com amor e coragem, embarca comigo em cada nova aventura. Compartilhar as alegrias e enfrentar os desafios diários ao seu lado tornou tudo mais leve. Sua presença constante, sua paciência nos momentos mais difíceis e seu apoio incondicional foram meu porto seguro em cada etapa desta jornada.

Dedico aos meus pais, Luís e Neide, exímios professores da educação pública brasileira, que sempre me inspiraram a trilhar o caminho da educação como fonte de transformação e crescimento. O cuidado, a sabedoria e o amor de vocês foram meu acalento ao longo de todos estes anos. Ao meu irmão Fernando e às minhas sobrinhas, Maria Fernanda e Maria Clara, por proporcionarem momentos de leveza e descontração em meio a esta fase.

Ao meu orientador, Prof. Dr. Pedro Menezes, por sua paciência, sabedoria e confiança. Sua orientação foi o que me iluminou nos momentos mais difíceis, dando-me segurança para avançar. Sou profundamente grata por acreditar no meu potencial e me encorajar a assumir o protagonismo das minhas escolhas tanto acadêmicas quanto profissionais.

Aos meus colegas da pós-graduação, que dividiram comigo os mesmos anseios, alegrias e desafios, oferecendo apoio e incentivo. Aos colegas de trabalho e do grupo de pesquisa, que tantas vezes me ajudaram com generosidade, apoio ou palavras encorajadoras.

Aos pontos focais das instituições que toparam contribuir com a pesquisa e se tornaram verdadeiros colegas de vida. A confiança e contribuição foram indispensáveis para a concretização deste estudo, reafirmando a força e a relevância do esforço coletivo na busca por soluções e melhorias.

A todos vocês, minha eterna gratidão.

Resumo

A Gestão de Desempenho (GD) é reconhecida como um sistema capaz de impulsionar resultados organizacionais superiores por meio do gerenciamento contínuo do desempenho individual. Contudo, estudos relatam a insatisfação por parte dos praticantes dada a dificuldade em identificar os resultados concretos e os impactos gerados pelo sistema de GD nas organizações. Esse cenário motivou diversas investigações sobre a efetividade do sistema, mas a mensuração desse construto tem se mostrado desafiadora devido à sua natureza multidimensional. Como consequência, a literatura apresenta estudos fragmentados, que não se integram, dificultando a consolidação de métodos para avaliar as contribuições e os resultados da GD nas organizações. Assim, foi proposto um modelo teórico-integrativo constituído a partir de revisão sistemática de literatura de 58 artigos. Este estudo, de abordagem mista, teve como objetivo testar empiricamente um modelo integrativo de avaliação da efetividade do sistema de GD no contexto da Administração Pública brasileira. Diante da profusão conceitual e das medidas difusas, foi realizado um estudo para desenvolver e apresentar evidências de validade para as escalas de percepção de Reação, Aprendizagem e Transferência. Foram realizadas 11 entrevistas com 14 praticantes de GD, seguidas pela validação das escalas com nove juízes especialistas no tema. Em uma segunda etapa, uma Análise Fatorial Exploratória (AFE) foi conduzida com 238 servidores públicos. Posteriormente, as escalas foram submetidas à Análise Fatorial Confirmatória (AFC) e realizado um teste de hipóteses com os dados coletados de 145 servidores vinculados a duas organizações públicas federais e que já participaram do processo de GD. Os resultados permitiram testar as relações entre as variáveis individuais de Reação, Aprendizagem e Transferência no modelo proposto e revelaram que não há relação direta entre elas, sugerindo que os elos podem depender de variáveis adicionais. Esta tese contribui para a literatura ao construir e validar três instrumentos de medidas individuais para mensurar a efetividade do sistema de GD. Além disso, propõe uma visão integrativa para o campo, fomentando uma agenda de pesquisa voltada à compreensão mais ampla das relações que sustentam a efetividade da GD. Por fim, oferece subsídios gerenciais, possibilitando avaliar como as pessoas pensam, aprendem e transferem comportamentos para o trabalho a partir de suas interações com o sistema de GD.

Palavras-chave: Gestão de Desempenho. Avaliação de Desempenho. Avaliação de Efetividade. Reação. Aprendizagem. Transferência. Setor Público. Modelo integrativo.

Abstract

Performance Management (PM) is recognized as a system capable of driving superior organizational results through the continuous management of individual performance. However, studies report dissatisfaction among practitioners due to the difficulty in identifying the tangible outcomes and impacts generated by PM systems in organizations. This scenario has motivated numerous investigations into the effectiveness of the system, but measuring this construct has proven challenging due to its multidimensional nature. Consequently, the literature presents fragmented studies that lack integration, complicating the consolidation of methods to evaluate the contributions and outcomes of PM systems in organizations. Thus, a theoretical-integrative model was proposed, developed from a systematic literature review of 58 articles. This mixed-methods study aimed to empirically test an integrative model for assessing the effectiveness of PM systems in the context of Brazilian Public Administration. In light of the conceptual proliferation and diffuse measures, a study was conducted to develop and provide validity evidence for scales measuring perceptions of Reaction, Learning, and Transfer. Eleven interviews were conducted with 14 PM practitioners, followed by the validation of the scales by nine subject matter experts. In a subsequent phase, an Exploratory Factor Analysis (EFA) was conducted with 238 public servants. The scales were then subjected to Confirmatory Factor Analysis (CFA), and a hypothesis test was performed using data collected from 145 public servants affiliated with two federal public organizations who had already participated in the PM process. The results allowed for testing the relationships among the individual variables, revealing no direct relationship among Reaction, Learning and Transfer. This finding suggests that the connections may depend on additional variables. This study contributes to the literature by developing and validating three individual-level measurement instruments to assess the effectiveness of PM systems. Furthermore, it proposes an integrative perspective for the field, fostering a research agenda aimed at a broader understanding of the relationships underpinning PM effectiveness. Finally, it offers managerial insights, enabling the evaluation of how individuals think, learn, and transfer behaviors to their work based on their interactions with the PM system.

Keywords: Performance Management. Performance Appraisal. Evaluation. Effectiveness. Reaction. Learning. Transfer. Public Sector. Integrative model.

LISTA DE FIGURAS

Figura 1. O Ciclo de Gestão de Desempenho.....	31
Figura 2. Avaliação de efetividade	36
Figura 3. Modelo de critérios avaliativos da efetividade de Gestão de Desempenho	38
Figura 4. Diagrama utilizado para revisão sistemática de literatura.....	41
Figura 5. Modelo teórico-integrativo de avaliação de efetividade do sistema de GD.....	46
Figura 6. Fluxo de atividades metodológicas do Estudo 1	101
Figura 7. Ocorrência das palavras e conexão entre elas	123
Figura 8. Nuvem de palavras	124
Figura 9. Dendograma das classes de palavras.....	125

LISTA DE TABELAS

Tabela 1	43
Tabela 2	47
Tabela 3	51
Tabela 4	53
Tabela 5	62
Tabela 6	64
Tabela 7	69
Tabela 8	71
Tabela 9	77
Tabela 10	78
Tabela 11	82
Tabela 12	83
Tabela 13	85
Tabela 14	85
Tabela 15	100
Tabela 16	102
Tabela 17	104
Tabela 18	106
Tabela 19	112
Tabela 20	118
Tabela 21	126
Tabela 22	127
Tabela 23	129
Tabela 24	136
Tabela 25	142
Tabela 26	147
Tabela 27	151
Tabela 28	152
Tabela 29	157
Tabela 30	160
Tabela 31	164
Tabela 32	168

Tabela 33	170
Tabela 34	171
Tabela 35	173
Tabela 36	176
Tabela 37	176
Tabela 38	178
Tabela 39	179
Tabela 40	181
Tabela 41	181
Tabela 42	184
Tabela 43	185
Tabela 44	185
Tabela 45	186
Tabela 46	187
Tabela 47	188
Tabela 48	188
Tabela 49	189
Tabela 50	190
Tabela 51	190

LISTA DE SIGLAS

AD	Avaliação de Desempenho
AFC	Análise Fatorial Confirmatória
AFC	Análise Fatorial Confirmatória
AFE	Análise Fatorial Exploratória
AFE	Análise Fatorial Exploratória
AMO	<i>Ability, motivation, opportunity</i>
APF	Administração Pública federal
AVE	Variância Média Extraída
BIC	<i>Bayesian Information Criterion</i>
CFI	<i>Comparative Fit Index</i>
CHD	Classificação Hierárquica Descendente
CO	Comportamento Organizacional
CVC	Coeficiente de Validade de Conteúdo
CVF	<i>Competing Values Framework</i>
ECV	<i>Explained Common Variance</i>
EPA-GD	Escala de Percepção de Aprendizagem da Gestão de Desempenho
EPR-GD	Escala de Percepção de Reação da Gestão de Desempenho
EPT-GD	Escala de Percepção de Transferência da Gestão de Desempenho
GD	Gestão de Desempenho
GEP	Gestão Estratégica de Pessoas
GP	Gestão de Pessoas
KMO	Kaiser-Meyer-Olkin
KSA	Conhecimento, habilidades e atitudes
MIREAL	<i>Mean of Item Residual Absolute Loadings</i>
ML	<i>Maximum Likelihood</i>
NPM	<i>New Public Management</i>
PAF	<i>Principal Axis Factoring</i>
PEC	Proposta de Emenda à Constituição
PGD	Programa de Gestão e Desempenho
PMSE	<i>Perceived performance management system effectiveness</i>
PRISMA	<i>Preferred Reporting Items for Systematic reviews and Meta-Analyses</i>
RDWLS	<i>Robust Diagonally Weighted Least Squares</i>

RMSEA	<i>Root Mean Square Error of Approximation</i>
SET	<i>Social exchange theory</i>
SRMR	<i>Standardized Root Mean Square Residual</i>
STs	Segmentos de texto
TCJ	Taxa de Concordância de Juízes
TCLE	Termo de Consentimento Livre e Esclarecido
TLI	<i>Tucker-Lewis Index</i>
UniCo	<i>Unidimensional Congruence</i>
VBR	Visão baseada em recursos
VD	Variáveis Critério
VI	Variáveis Predictoras

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Problematização	15
1.2	Objetivos	17
1.3	Justificativas.....	18
2	REFERENCIAL TEÓRICO	20
2.1	Gestão Estratégica de Pessoas	20
2.1.1	Efetividade das Políticas e Práticas de Gestão Estratégica de Pessoas....	25
2.2	Gestão de Desempenho.....	30
2.2.1	Efetividade de Gestão de Desempenho.....	35
2.3	Proposição de um Modelo Teórico-Integrativo de Avaliação da efetividade do sistema de Gestão de Desempenho	39
2.3.1	Medidas do Modelo Teórico-Integrativo de Avaliação de Efetividade do Sistema de Gestão de Desempenho	47
2.4	Relações Empíricas de Avaliação de Efetividade do Sistema de Gestão de Desempenho.....	87
2.4.1	Reação → Aprendizagem	88
2.4.2	Aprendizagem → Transferência.....	92
2.4.3	Transferência → Resultados Operacionais.....	93
2.4.4	Resultados Operacionais → Resultados Financeiros.....	93
2.4.5	Resultados Financeiros → Impacto Societal	95
3	MÉTODOS, TÉCNICAS E PROCEDIMENTOS DE PESQUISA	99
3.1	Características da pesquisa	99
3.2	Percurso Metodológico do Estudo 1	101
3.3	Percurso Metodológico do Estudo 2	119
4	RESULTADOS	121
	Resultados do Estudo 1	121
4.1.	Fase A: Construção teórica e validação de juízes das Escalas de Percepção de Reação, de Aprendizagem e de Transferência.....	121
4.1.1.	Processamento dos Dados Textuais	123
4.1.2.	Análise de Conteúdo dos Dados Semiestruturados.....	127
4.1.3.	Análise de Conteúdo dos Dados Estruturados	145

4.1.4. Validação de juízes	150
4.2. Fase B: Validação empírica	167
4.2.1. Subfase 1B: Análise Fatorial Exploratória (AFE)	167
4.2.2. Subfase 2B: Análise Fatorial Confirmatória (AFC)	183
Resultados do Estudo 2	189
4.3. Teste de Hipóteses Individuais	189
5. CONCLUSÃO	194
REFERÊNCIAS	200
Apêndice A - Carta de Apresentação da Entrevista (e-mail)	220
Apêndice B – Instrumento 1 – Roteiro de Entrevista	221
Apêndice C – Instrumento 2 – Validação de juízes	232
Apêndice D – Questionário AFE	235

1 INTRODUÇÃO

Avaliar o desempenho das pessoas no trabalho tem sido de interesse de teóricos e praticantes há pelo menos 100 anos (DeNisi & Murphy, 2017). Ao longo desse período, os estudos se desenvolveram durante grande parte do tempo em torno da temática da Avaliação de Desempenho (AD) até o reconhecimento de que gerenciar estrategicamente o desempenho das pessoas era essencial para o desenvolvimento e a sobrevivências das organizações (Brown et al., 2019; Den Hartog et al., 2004).

Esse movimento advém das pesquisas iniciadas durante a década de 1980, que passam a reconhecer os recursos humanos como importante fonte de vantagem competitiva das organizações (Guest, 1987, 1997; Jackson & Schuler, 1995b). A descoberta acontece diante do cenário de organizações cada vez mais competitivas decorrente do mercado internacional globalizado, fazendo com que as fontes tradicionais de vantagem competitiva fossem rapidamente absorvidas pelas organizações, momento em que teóricos e praticantes compreendem que o gerenciamento adequado da força de trabalho organizacional poderia atender às novas demandas do mercado e promover vantagem competitiva (Bell et al., 1983).

Diante disso, tem-se uma mudança na função de Gestão de Pessoas (GP) nas organizações, que é ampliada para além daquela tradicional maneira de administrar os recursos humanos e que agora seria capaz de produzir e avaliar seus resultados, assumindo um papel estratégico na organização (Barney & Wright, 1998; Guest, 1997; Jiang & Messersmith, 2018a; Kuipers & Giurge, 2016). Para isso, fala-se na necessidade de alinhamento das políticas e práticas de GP com a estratégia organizacional a fim de promover resultados organizacionais superiores, originando, assim, o campo de estudos da Gestão Estratégica de Pessoas (GEP).

Em consonância com esses princípios, têm-se a Gestão de Desempenho (GD) como um importante sistema estruturante das decisões estratégicas entre GP e a organização, responsável por identificar, mensurar e desenvolver o desempenho dos indivíduos em um processo contínuo alinhado à estratégia organizacional a fim de promover desempenho organizacional elevado (Aguinis, 2013).

Desde então, o sistema de GD está presente em grande parte das organizações ao redor do mundo para mensurar o desempenho de seus funcionários, promovendo feedbacks contínuos e políticas de recompensa, promoção ou desenvolvimento (*e.g.* Cascio, 2015; Lawler, 2003; Levy & Williams, 2004). Para além das organizações privadas, as organizações públicas de diversos governos internacionais também têm adotado a GD como parte da gestão estratégica

de pessoas (e.g. Abu-Shanab et al., 2021; Hondegheem & Dorpe, 2013; Longenecker & Nykodym, 1996; Sotirakou & Zeppou, 2006).

No Brasil, a GD também é adotada na Administração Pública em decorrência do processo de reformas gerenciais empreendidas a partir do final do século XX (Coelho et al., 2016; Martins & Ensslin, 2020; Odelius & Santos, 2007; Vilhena & Martins, 2022). Assim, tem-se uma série de mudanças a partir da adoção de um modelo pós-burocrático de gestão pública que se baseia em valores profissionalizantes e de competitividade inspirados pelo movimento da Nova Administração Pública (ou *New Public Management* – NPM), que prevê a adoção de políticas, práticas e valores gerenciais típicos do setor privado no setor público (Barzelay & Jacobsen, 2009; Hood, 1995; Secchi, 2009).

Dessa forma, a mensuração do desempenho no setor público brasileiro tem seu início ainda na década de 1970, mas a base normativa que regula os mecanismos da AD é promulgada em 2008 pela Lei nº 11.784/08, sendo ampliada com a edição do Decreto nº 7.133, em 2010, que confere espaço para que a GD na Administração Pública fosse instituída nas organizações (Elvira, 2018). Recentemente, o tema de GD voltou a fazer parte das discussões sobre a reforma administrativa com a Proposta de Emenda Constitucional nº 32/2020, que pode conferir peso ainda maior aos sistemas de GD nas organizações públicas em termos da regulamentação do desligamento de servidores públicos com desempenho insuficiente (e.g. Neiva, 2020; Secchi et al., 2021; Vilhena & Martins, 2022).

Logo, é interessante perceber que a GD tem sido adotada formalmente pela maioria das grandes organizações, sejam elas públicas ou privadas, demonstrando que o sistema continua atual e recorrente tanto na literatura quanto na prática. Isto tem acontecido em meio a diversos apontamentos por parte de estudos recentes acerca de reclamações e da insatisfação em relação à GD advindas dos funcionários (DeNisi & Murphy, 2017; Haines & St-Onge, 2012; Levy et al., 2017; Kevin R. Murphy, 2020; Schleicher et al., 2018).

Mesmo com a presença de GD na literatura científica e na prática, principalmente nos últimos anos e com toda explanação teórica feita pelos estudos, o sistema de GD têm sido operados distantes da perfeição. A modelagem do sistema de GD e a diferenciação conceitual estabelecida entre AD e GD com intuito de cravar o espaço e a importância ocupada por cada uma, na prática, não têm evitado que ambas acabem sendo vistas como falhas por funcionários e gestores (e.g. Ikramullah et al., 2016; Murphy, 2020; Schleicher et al., 2018).

1.1 Problematização

A insatisfação com a forma de gerenciamento do desempenho individual é conhecida há algum tempo por teóricos, mas principalmente pelos praticantes; nos últimos anos, tem crescido a preocupação de acadêmicos de que apenas ajustes técnicos ou metodológicos no sistema de GD não são suficientes para alterar a forma com que as pessoas o percebem (Levy et al., 2017).

Assim, o sistema de GD permanece sendo uma prática cujos resultados se mostram distantes sob o olhar dos atores organizacionais (Cappelli & Conyon, 2017). Diante disso, acadêmicos e praticantes passam a questionar se todo investimento feito para implementar e promover a GD nas organizações realmente traz benefícios substanciais à organização (Awan et al., 2020; Cappelli & Conyon, 2017; Dewettinck, 2008; Dewettinck & van Dijk, 2013; Ikramullah et al., 2016; Iqbal et al., 2015, 2019; Kakkar et al., 2020; Keeping & Levy, 2000; Lawler, 2003; Modipane et al., 2019; K. R. Murphy & Cleveland, 1995; Pichler, 2012; Schleicher et al., 2018, 2019; Sharma et al., 2016; Tuytens & Devos, 2012), conforme alguns exemplos:

Revisões de pesquisas sobre AD e GD observaram que há pouca ou nenhuma evidência de que esses sistemas tenham algum impacto real no desempenho do funcionário (Murphy, 2020, p. 3).

Apesar da onipresença de GD na prática e na literatura de pesquisa, ainda permanece sem resposta a questão da sua efetividade (Schleicher et al., 2018, p. 2.210).

O problema é que poucos estudos apoiam as muitas afirmações sobre as contribuições reais de várias práticas para a efetividade geral dos sistemas de gestão de desempenho (Haines & St-Onge, 2012, p. 1.171).

Esses questionamentos de GD são um reflexo do que também acontece com a produção de GEP, a qual tem sido submetida, desde a sua concepção, a diversos exames que buscam demonstrar o que envolve a relação entre GP e desempenho (Guest, 1997). Porém os estudos não olham todos para o mesmo objeto, fornecendo, em uma perspectiva geral, achados e conclusões cumulativos que não auxiliam a compreender os resultados de GP para a

organização. Após mais de trinta anos de produção do campo de GEP, teóricos afirmam que o tema apresenta tendências controversas, definições incertas, relações mínimas entre GP e desempenho organizacional e pouco rigor metodológico (Cooke et al., 2021; Jackson et al., 2014; Jiang & Li, 2019; B. E. Kaufman, 2015a, 2020; Lengnick-Hall et al., 2009; Paauwe, 2011; Paauwe & Boselie, 2005; Schuler & E. Jackson, 2014).

Tais apontamentos permanecem sem resposta, uma vez que demonstrar a efetividade, em específico, de sistema de GP mostra-se uma tarefa complexa e de difícil execução, mais até do que a de acessar a efetividade de outras atividades organizacionais (Ahmed, 1999). A efetividade dos sistemas de GD envolve lidar com a variável de desempenho, sobretudo o organizacional, o qual possui uma característica multidimensional que permite com que ele seja avaliado de diferentes formas, dificultando, assim, a sua mensuração (Guest, 1997; Paauwe, 2009). Diante disso, nos últimos anos as pesquisas têm buscado investigar a efetividade dos sistemas de GD utilizando certos critérios de mensuração para acessar as relações antecedentes- resultados que manifestam a efetividade de GD (Iqbal et al., 2015).

Entretanto, o que há é um acúmulo de estudos que não se integram, pois são por muitas vezes adotados de forma isolada e distantes de uma abordagem holística (Ikramullah et al., 2016; Iqbal et al., 2019; Levy & Williams, 2004a; Schleicher et al., 2019; Zahra & Jadoon, 2016). Isso prejudica o desenvolvimento do campo científico e prático em razão da ausência de teorias e métodos objetivos e evidentes, cujos critérios realmente forneçam contribuições na identificação da efetividade do sistema de GD; e mais do que isso: a falta de uma sistematização dos estudos acerca dos critérios de efetividade de GD dificulta o conhecimento cumulativo do campo sobre os critérios relevantes e como eles podem contribuir para identificação dos resultados relacionados ao desempenho organizacional.

A ausência da sistematização de critérios reflete na operacionalização metodológica dos estudos que, em razão da dificuldade de clareza entre os critérios, acabam adotando medidas extremamente genéricas e impedem de tirar conclusões claras acerca da efetividade do sistema de GD (Schleicher et al., 2019). A necessidade de maior clareza na definição constitutiva de alguns critérios a fim de sanar confusões conceituais foi apontada ainda em 1995 por Murphy e Cleveland, mas persiste conforme apontado por Schleicher et al. (2019).

Logo, havendo essa profusão de conceitos e medidas que dificultam a mensuração de tais dimensões, o questionamento acerca da capacidade dos sistemas de GD enquanto ferramenta para promoção de comportamentos individuais superiores, capazes de promover desempenho organizacional, permanece. Sobretudo no caso do setor público brasileiro, destaca-

se que diante das dificuldades da GD nas organizações e das recentes discussões lançadas sobre a aplicação do sistema nas organizações públicas, a identificação da efetividade de GD mostra-se saliente.

É em razão desse contexto que Kaufman (2020, p. 66) resume a mensuração dos resultados de sistema de GEP, como visto no caso da GD, como “centrada em um paradigma improdutivo quase universalista, orientado por métodos quantitativos, com conceitos do sistema de GP mal elaborados e medidos e um conjunto esfarrapado de implicações práticas e resultados”. Diante de tal panorama, Budd (2020) e Kaufman (2020) ressaltam a importância de voltar a atenção dos resultados de GP para uma perspectiva menos universalista e cuja produção foque no campo teórico de desenvolvimento de teorias específicas para GP, como feito pelos fundadores do campo da GEP (*e.g.* Huselid, 1995; Lawler, 1986; Wright et al., 1992), os quais entendiam a GP como sistemas abertos, de gestão generalista, configuracional, com interesses plurais, abrangendo, ainda, experiências práticas e profissionais dos próprios pesquisadores do campo.

Portanto, destaca-se a necessidade de estudos que respondam aos anseios dos acadêmicos, fornecendo uma nova abordagem integrativa, alinhada para além da tendência universalista do campo, que amplie o desenvolvimento das teorias voltadas para o campo específico de GEP; e dos praticantes, com insumos sobre os resultados do sistema de GD que atenda aos interesses das unidades de GP nas organizações públicas sob uma abordagem menos generalista e aplicada ao setor público.

1.2 Objetivos

O objetivo geral deste estudo consiste em testar empiricamente um modelo de avaliação de efetividade do sistema de Gestão de Desempenho para o contexto da Administração Pública Federal brasileira.

Objetivos específicos:

1. Identificar dimensões utilizadas para avaliar a efetividade dos sistemas de Gestão de Desempenho;
2. Desenvolver um modelo integrativo de avaliação de efetividade dos sistemas de Gestão de Desempenho para o contexto da Administração Pública brasileira;

3. Demonstrar empiricamente as relações entre as dimensões utilizadas para avaliar a efetividade do sistema de Gestão de Desempenho no contexto da Administração Pública brasileira.

1.3 Justificativas

Inicialmente, este estudo mostra-se relevante na perspectiva teórica ao sistematizar a produção sobre avaliação de efetividade de GD, identificando as dimensões e critérios adotados pela literatura do campo. A partir disso, é possível responder às demandas dos autores sobre a descrição das possibilidades de avaliação, delineando aquelas mais consolidadas e preteridas pela literatura de GD, promovendo um modelo de avaliação de efetividade de GD mais enxuto e teoricamente robusto.

O desenvolvimento de um modelo integrativo composto pelas dimensões e os critérios adotados pela literatura propõe uma abordagem holística e complementar para a avaliação de efetividade de GD. Esse modelo permite visualizar a posição das dimensões adotadas pela literatura em relação ao sistema de GD, assim como a identificação das relações entre elas, a qual possibilita que haja uma demonstração empírica da GD com o desempenho organizacional, capaz de determinar a efetividade do sistema de GD.

Outra contribuição consiste na realização da pesquisa no setor público, favorecendo a construção de um modelo de avaliação de efetividade para a realidade da Administração Pública brasileira, ampliando as possibilidades de discussão em meio ao tema ainda pouco estudado pela literatura nesse contexto.

Com relação aos aspectos metodológicos, este estudo se destaca por fornecer evidências empíricas das relações entre as variáveis individuais do modelo integrativo de avaliação da efetividade da GD. Os resultados contribuem para a compreensão de como esses elementos se conectam dentro do modelo proposto, fornecendo subsídios que podem, em estudos futuros, explorar a relação entre o sistema de GD e o desempenho organizacional. Esse modelo permite a formulação de hipóteses de pesquisa e o teste das relações entre variáveis de perspectiva individual, além de possibilitar a construção e validação de medidas de forma exploratória e confirmatória, adaptadas ao contexto do setor público brasileiro.

Espera-se, assim, superar confusões conceituais diante da profusão de conceitos dúbios e medidas aplicadas de forma genérica que impedem determinadas conclusões sobre a efetividade de GD.

Por último, destaca-se que este estudo tem relevância prática e gerencial ao fornecer um modelo de avaliação da efetividade de GD aplicado ao contexto das organizações públicas, útil para orientar o desenvolvimento de sistemas de GD mais robustos e eficazes. O teste empírico do modelo permite verificar, inicialmente, se o sistema de GD gera benefícios para os indivíduos, já que a identificação de mudanças no comportamento individual é uma premissa central para que o sistema alcance, eventualmente, impactos organizacionais.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta fundamentos e perspectivas teóricas acerca da avaliação de efetividade do sistema de Gestão de Desempenho (GD). Para isso, a primeira seção apresenta um retrospecto teórico sobre o campo da Gestão Estratégica de Pessoas (GEP), desde a fundação, passando pelos seus modelos e complementando com estudos sobre a avaliação de efetividade das políticas e práticas de GP e sua relação com o desempenho organizacional. Na segunda seção, é explorado o conceito de Gestão de Desempenho (GD), enfatizando seu papel como ferramenta estratégica no contexto da GEP e as questões relacionadas à sua efetividade. Em seguida, a terceira seção propõe um modelo teórico-integrativo de avaliação da efetividade do sistema de GD, consolidando as dimensões e critérios identificados na literatura revisada. Por fim, na quarta seção, são discutidas as relações entre as dimensões do modelo de avaliação da efetividade do sistema de GD, estabelecendo as bases conceituais e empíricas para os objetivos desta pesquisa.

2.1 Gestão Estratégica de Pessoas

Na década de 1980, o novo mercado internacional globalizado promoveu um contexto organizacional altamente competitivo cujas fontes de vantagens tradicionais (*e.g.* tecnologias, qualidade, estratégias de vendas etc.) foram rapidamente absorvidas pelas organizações até que se tornarem facilmente imitáveis (Kaplan & Norton, 1996). A partir disso, os praticantes se voltaram para o recurso organizacional “pessoas” ao perceberem que o gerenciamento adequado desta força de trabalho poderia atender às novas demandas do mercado e promover vantagem competitiva (Bell et al., 1983).

Segundo Wright e Ulrich (2017), essas novas demandas trouxeram a necessidade de mudanças no enfoque dos efeitos das práticas de GP. Inicialmente voltado exclusivamente para o nível individual, o debate passou a incluir também impactos em nível organizacional. Além disso, tornou-se fundamental considerar os sistemas de GP como um todo, em vez de analisar apenas práticas isoladas. Por fim, os autores destacam a importância de discutir a gestão de recursos humanos no contexto mais amplo da estratégia organizacional, promovendo um alinhamento entre pessoas, processos e objetivos organizacionais.

Diante desse contexto, os estudos destacam a importância de alinhar as políticas e práticas de GP com a estratégia organizacional, premissa que fundamenta o campo da Gestão Estratégica de Pessoas (GEP) (Wright & McMahan, 1992), contexto em que os recursos humanos figuraram como uma importante fonte de vantagem competitiva (Guest, 1987, 1997; Jackson & Schuler, 1995b). Nesse contexto, a atuação de Recursos Humanos nas organizações deixa de se restringir à administração tradicional e passa a assumir uma função estratégica, focada na geração e avaliação de resultado, conferindo à área uma posição central na estratégia organizacional (Barney & Wright, 1998; Guest, 1997; Jiang & Messersmith, 2018b; Kuipers & Giurge, 2016).

A partir do reconhecimento do caráter estratégico da GP, a literatura sobre GEP avançou significativamente nas últimas décadas, consolidando diversas definições (*e.g.* Boxall & Purcell, 2000; Guest, 1997, 2011; Jackson et al., 2014; Lepak et al., 2006; Pauwee, 2009; Wright & McMahan, 1992). Entre essas, destaca-se a definição de Jackson, Schuler e Jiang (2014, p. 2), que conceituam a GEP como um campo de estudos voltado para os sistemas de GP e suas inter-relações com outros elementos organizacionais. Esses elementos incluem o ambiente interno e externo da organização, os diferentes *stakeholders* que interagem com os sistemas de GP, bem como aqueles que avaliam a efetividade organizacional e influenciam sua sobrevivência a longo prazo.

Além da preocupação com a definição conceitual de GEP, estudos e revisões ressaltam que a principal motivação do campo de GEP tem sido demonstrar essas inter-relações entre os sistemas de GP e o desempenho organizacional (Guest, 1997; Jackson et al., 2014; Paauwe & Boselie, 2005). Para isso, os estudos adotam uma série de teorias a fim de explicar o porquê, como e quando os sistemas de GP estão relacionados aos resultados organizacionais para que esses sistemas se mostrem capazes de promover desempenho organizacional superior (Jiang & Messersmith, 2018). Ainda em 1992, Wright e McMahan já haviam identificado seis teorias predominantes que foram aplicadas para compreender essa relação. Porém, destas, Wright e Ulrich (2017) destacam em específico três paradigmas teóricos que dominam o campo de pesquisa da área: a visão baseada em recursos (VRB, *resource-based view* – RBV), teoria do capital humano e teoria de troca social (*social exchange theory* – SET). Outros paradigmas teóricos como a perspectiva comportamental e a ferramenta de habilidade-motivação-oportunidade (*ability-motivation-opportunity* – AMO) são citados pelos autores em razão da frequência com que aparecem na literatura de GEP (Cooke et al., 2021; Delery & Roumpi, 2017; Jiang & Messersmith, 2018a; Paauwe & Boselie, 2005).

O potencial dos recursos humanos em promover vantagem competitiva é apoiado pela visão baseada em recursos (VBR) (ou *resource-based view* – RBV) (e.g. Jackson & Schuler, 1995a; Snell et al., 1996; Patrick M. Wright et al., 1994) pois, segundo essa visão, para obter a diferenciação e a vantagem competitiva desejada, os recursos organizacionais, tais como o capital humano, tecnológicos, financeiro e outros, deveriam ser vistos como forças da organização (Barney, 1991). Nesse sentido, o capital humano figura como um recurso capaz de agregar valor para a organização, raro, de difícil imitação e insubstituível (Barney & Wright, 1998). É a partir dessa visão que os estudos passam a discutir sobre de que forma esses quatro critérios podem ser trabalhados pela GEP (Jiang & Messersmith, 2018).

Essa é uma das primeiras teorias a fornecer uma explicação convincente acerca da relação dos recursos humanos com os resultados organizacionais e, em razão disso, utilizada por quase todo estudo sobre a temática de GEP (Kaufman, 2015; Wright & Ulrich, 2017).

Embora a VRB ofereça uma base sólida para justificar o potencial estratégico dos recursos humanos, críticas apontam que a teoria não fornece uma explicação clara de como as práticas de GP entregam resultados organizacionais concretos. Essa visão tem recebido uma série de questionamentos e críticas (Jiang & Messersmith, 2018b; Wright & Ulrich, 2017).

Ainda em 2000, Boxall e Purcell (2000) alertaram sobre a VBR permitir convencer, em certa medida, sobre a relação entre GP e vantagem competitiva, porém não respondia sobre como ela de fato poderia ser usada. Em 2006, Lepak et al. (2006) ainda identificaram a mesma dificuldade, alertando sobre os estudos continuarem não explorando profundamente como GP pode, de fato, entregar os quatro critérios previstos pela VBR. Roumpi e Delery (2019) complementam que garantir que os quatro critérios de vantagem competitiva sejam atingidos pelos sistemas de GP é pouco provável utilizando como exemplo a “inimitabilidade”, pois ao longo do tempo outras organizações podem replicar determinadas práticas e sistemas de GP. Kaufman (2015) também identifica alguns problemas relacionados à adoção da VBR na literatura de GEP ao utilizar a ótica econômica para avaliar o uso da VBR na pesquisa de GEP.

A teoria do capital humano, amplamente adotada na literatura de GEP, explora como investimentos no desenvolvimento de características individuais valiosas para as organizações podem gerar vantagem competitiva. Segundo Burke e Cooper (2007), práticas de GP que promovem conhecimento, habilidades e atitudes (KSAs) contribuem para a criação de um capital humano valioso, raro, inimitável e insubstituível, com maior potencial de impactar positivamente o desempenho organizacional.

Diferentemente de outros recursos organizacionais, o capital humano é detido pelo próprio funcionário/indivíduo, logo, caso ele deixe a organização, a vantagem competitiva pode ser transferida para outra organização (Wright et al., 2001). Diante disso, se cria uma discussão sobre o que, de fato, representaria fonte de vantagem de competitiva: os funcionários, em termo de conhecimentos e habilidades valoráveis para a organização, ou as práticas de GP, que podem promover determinado nível de capital humano ou prevenir a sua perda (Boselie, Van Harten, et al., 2019; Jiang & Messersmith, 2018a). Wright et al. (2001) destacam um aspecto importante que auxilia a entender a relação entre os sistemas de GP como meios pelos quais a organização pode continuar gerando vantagem ao longo do tempo a partir de funcionários que entram e saem, o que alterna os comportamentos desejados ao longo do tempo atendendo a mudanças organizacionais e contingências estratégicas.

Se a VBR e a teoria do capital humano explicam como os recursos humanos e as práticas de GP são importantes, a perspectiva comportamental enfatiza como o sistema de GP pode auxiliar a alcançar resultados organizacionais (Jiang & Messersmith, 2018). A perspectiva comportamental permite compreender como o ambiente interno (*e.g.* estratégia organizacional) e externo (*e.g.* novas tecnologias) determinam o papel comportamental que deve ser considerado no desenho de sistemas de GP (Jackson & Schuler, 1996). A perspectiva comportamental fornece uma base teórica para examinar os mecanismos que mediam a relação entre os sistemas de GP e os resultados organizacionais (Jiang & Messersmith, 2018).

Uma variante da perspectiva comportamental é a teoria AMO, sigla em inglês para indicar três componentes do desempenho individual: A – *ability* (habilidade), M – *motivation* (motivação) e O – *opportunity* (oportunidade), utilizados para compreender como as práticas de GP fornecem impacto nos resultados organizacionais (Jackson et al., 2014). Essa teoria tem sido recorrentemente adotada pela literatura de GEP com intuito de fornecer insumos que expliquem o processo de mediação entre as práticas de GP e o desempenho organizacional (*e.g.* Guest, 1997).

A teoria AMO propõe que práticas de Gestão de Pessoas (GP) geram resultados organizacionais ao desenvolver habilidades (*ability-enhancing*), motivar funcionários (*motivation-enhancing*) e criar oportunidades para que contribuam efetivamente (*opportunity-enhancing*). Essas práticas fortalecem o capital humano e alinham os esforços dos funcionários aos objetivos organizacionais, conectando GP ao desempenho organizacional (Roumpi & Delery, 2019).

À medida que os pesquisadores passaram a se concentrar nos fatores que influenciam a relação entre as práticas de GP e o desempenho organizacional, com foco em atitudes e comportamentos dos funcionários, a teoria de troca social (*social exchange theory* – SET) recebeu maior ênfase na literatura de GEP (Jiang & Messersmith, 2014). A SET, com sua ênfase no nível individual, explora como interações justas entre empregadores e empregados promovem um sentimento de reciprocidade. Essa percepção leva os indivíduos a se engajarem em esforços adicionais, buscando manter um equilíbrio justo nas trocas entre ambas as partes (Kepes & Delery, 2006; Wright & Ulrich, 2017).

É importante destacar que até aqui foram reunidas e explicitadas aquelas teorias com maior frequência na literatura de GEP, porém existem tantas outras que fundamentam o campo e que também são adotadas (*e.g.* teoria institucional, teoria da agência, abordagens universalista, contingencialista e configuracional, entre outras) (cf. Cooke et al., 2021; Jiang & Li, 2019; Jiang & Messersmith, 2018a; P. M. Wright & McMahan, 1992; Patrick M. Wright & Ulrich, 2017).

De toda forma, há um consenso sobre as teorias adotadas pelos estudos que abordam GEP serem predominantemente teorias originárias do campo de Comportamento Organizacional (CO) em que mecanismos de mediação a nível individual predominam, demonstrando uma orientação psicológica do campo de pesquisa de GEP (Cooke et al., 2021). Kaufman (2015b) apontou, ainda em 2015, que a lente teórica utilizada pelos estudos de GEP é voltada, principalmente, para perspectiva micro e para aspectos organizacionais internos que enfatizam comportamentos individuais e trajetórias psicossociais.

As teorias foram adotadas pelos estudos de GEP com intuito de discorrerem sobre se e como as políticas e práticas de GP alinhadas à estratégia organizacional poderiam promover um desempenho organizacional superior e, assim, vantagem competitiva. Mas a adoção de teorias de CO deixou claro que a variável desempenho é de difícil mensuração, tendo em vista a sua característica multidimensional que permite com que seja avaliada de diferentes formas (Guest, 1997; Paauwe, 2009).

No caso dos estudos de GEP, tem-se o que é chamado de “problema de critério”, pois, segundo a literatura de psicologia social, para avaliar indivíduos, programas ou intervenções organizacionais, é necessário que se tenha critérios, os quais, no caso do desempenho, são de difícil conceituação e mensuração. Por essa razão, Guest (1997) afirma que há um movimento crescente para estudos buscando examinar essa relação entre GP e desempenho, examinando aspectos particulares, porém os estudos não olham todos para o mesmo objeto, fornecendo, em

uma perspectiva geral, achados e conclusões cumulativos que auxiliem a compreender os resultados de GP para a organização.

Essa dificuldade persiste, como apontado mais recentemente por Boon et al. (2024), que destacam a falta de progresso em pesquisas voltadas a identificar os efeitos dos antecedentes e consequentes das práticas estratégicas de GP ao longo do tempo. Da mesma forma, Ogbonnaya et al. (2023) enfatizam o interesse crescente nas últimas três décadas em compreender como as práticas de estratégias de pessoal impactam o desempenho organizacional. Ainda assim, ambos os estudos deixam claro que, apesar dos esforços, ainda não há uma relação consistente e robusta entre esses aspectos, evidenciando a continuidade do desafio teórico e metodológico na área.

Logo, embora a literatura tenha avançado significativamente na compreensão dos efeitos das práticas e sistemas de GP, a fragmentação dos estudos e a ausência de critérios claros para avaliar o desempenho organizacional dificultam uma visão consolidada. Essa complexidade resulta em estudos que, embora contribuam individualmente, ainda carecem de uma visão unificada que explique de forma abrangente os impactos das práticas de GP. Na seção seguinte será abordado como esses estudos buscam avaliar a relação existente entre políticas e práticas de GP e o desempenho organizacional.

2.1.1 Efetividade das Políticas e Práticas de Gestão Estratégica de Pessoas

O estudo dos resultados obtidos pela adoção de políticas e práticas de GP capazes de fornecer desempenho organizacional superior tem crescido desde o final da última década do século passado (*e.g.* Dyer & Reeves, 1995; Guest, 1997; Huselid, 1995; Ulrich, 1998). Nesse intento, os estudos buscam explorar diferentes critérios, aqui definidos segundo a literatura de pesquisas em psicologia como sendo “um indicativo ou uma amostra de desempenho, mensurado diretamente ou indiretamente, e percebido como de valor para os constituintes organizacionais por facilitar decisões” (Austin & Villanova, 1992, p. 838). A partir da mensuração desses critérios, os autores afirmam que é possível representar em partes relações causais como sendo representantes do desempenho de GP capazes de promover efetividade organizacional, o que indicaria a efetividade das políticas e práticas de GP.

Para que se possa compreender o que está contido quando se faz avaliação de efetividade de GP, utiliza-se a literatura específica de avaliação de políticas públicas, a qual também se

estende para a avaliação de políticas, programas e intervenções organizacionais (*e.g.* Austin; Villanova, 1992) em que a efetividade é definida como sendo “o quanto uma política atinge aos objetivos previamente propostos” (Nagel, 1986, p. 99). Este é o mesmo conceito adotado por Ahmed (1999) ao estudar a mensuração da efetividade de práticas de GP.

Tendo em vista que, em linhas gerais, as práticas estratégicas de GP são usualmente desenhadas para promover desempenho organizacional, a avaliação de efetividade de GP envolve, então, avaliar esse construto multidimensional, incluindo vários aspectos e cuja distância lógica entre si é apontada por ser grande, o que torna a mensuração problemática (Baird, 2017; Baird et al., 2022; Banks & Kepes, 2015; Boselie, Paauwe, et al., 2019; Den Hartog et al., 2004; Guest, 1997; Huselid, 1995; Paauwe, 2011).

Porém, para além do objetivo geral do sistema de GD, deve-se ter claro que também existem objetivos específicos que variam a depender das necessidades, valores e preferências de grupos de *stakeholders* (executor das atividades, gerente, diretor, acionista, consumidor, cidadão, sindicalista etc.), o que torna mais difícil de se ter um consenso sobre qual objetivo seria útil para determinar a efetividade do sistema de GD diante do conceito de efetividade (Colakoglu et al., 2006; Ikramullah et al., 2016). Logo, assume-se que não há um critério exclusivo ou que seja mais importante em razão dos motivos explicitados anteriormente, o que ressalta a necessidade de se investigar múltiplos critérios, considerando a interrelação existente entre eles, assim como a posição ocupada por cada um, sendo uns mais proximais e outros mais distantes em relação à prática de GP (Colakoglu et al., 2006; Lepak, 2007; Ikramullah et al., 2016).

Dessa forma, ao procurar pelo potencial de critérios de efetividade de GP, Dyer e Reeves (1995) afirmam que as políticas e práticas de GP primeiro afetam resultados proximais e imediatos (indivíduos, equipes e operacionalização), que irão afetar resultados distais e mediatos (financeiros e do mercado externo à organização). Assim, os autores pontuam três categorias defensáveis de se avaliar o desempenho da organização, os quais seguem esta ordem lógica de posição em relação à prática de GP: uma política ou prática de GP tem seu primeiro impacto em i) resultados da unidade de GP, como critérios de absenteísmo, rotatividade e desempenho individual ou de equipe, os quais são apontados por serem os que mais têm seus resultados impactados pelas práticas de GP, pois são estes que, inicialmente, representam o objetivo para o qual tal política ou prática de GP é desenhada; a segunda categoria, esta mais distal da prática de GP do que os resultados anteriores, refere-se aos ii) resultados organizacionais, como produtividade, qualidade e serviço; e a terceira categoria é a mais

distante da prática de GP e se refere aos iii) resultados financeiros ou contábeis, como retorno financeiro de investimento ou de ativos (Dyer & Reeves, 1995).

O estudo de Huselid (1995) se mostrou notório ao encontrar suporte em suas afirmações de que investimentos em práticas de pessoal estavam associadas com baixa rotatividade e aumento da produtividade e do desempenho financeiro organizacional. Já Delery e Doty (1996) adotaram indicadores financeiros como critério de desempenho e identificaram relação significativa com as práticas de GP financeiros em bancos. A adoção do desempenho financeiro como medida para mensuração dos resultados de GP foi apontada como uma tendência na década de 1990 (Ulrich, 1998). Segundo meta-análise realizada por Boselie et al. (2005), metade dos estudos analisados considerava indicadores financeiros, sendo que lucro é apontado como um dos mais utilizados. Entretanto, adoção exclusiva desse tipo de medida para precisar os resultados de GP foi apontada como problemática, tendo em vista que os indicadores financeiros são influenciados por diversos outros fatores que não envolvem somente a GP (Paauwe & Boselie, 2005).

Outros modelos que buscaram demonstrar a relação entre GP e desempenho organizacional também estabelecem uma ordem lógica semelhante em que as práticas de GP primeiro impactam nos resultados individuais (*e.g.* compromisso organizacional e motivação), o que interfere no desempenho individual e, posteriormente, no organizacional (Bell et al., 1983; Guest, 1997; Paauwe & Richardson, 1997).

É importante destacar que Den Hartog et al. (2004) demonstram suas preocupações ao visualizarem a tendência do campo à época de realizar o teste empírico de tais modelos considerando resultados decorrentes de diferentes níveis de análise (*e.g.* individual, equipe e organizacional), dispostos em uma ordem lógica estabelecida, normalmente, de forma direta, por meio de estudos de recorte temporal com coleta de dados em apenas um dado momento, sem técnicas estatísticas rigorosas, como impõem esses estudos, e sem considerar como as práticas de GP são percebidas pelos funcionários.

Nesse sentido, na última década, os autores de estudos mais recentes se voltaram para as percepções dos funcionários, as quais são relacionadas às experiências com as práticas de GP, porém, mesmo com essa inclusão, há um consenso na área de que algumas dessas dificuldades de mensuração dos resultados das políticas e práticas de GP ainda perduram (*e.g.* Cascio, 2015b; Cooke et al., 2021; Jackson et al., 2014).

A maioria dos estudos sobre formas com que a GEP pode contribuir para o alcance dos objetivos organizacionais continuam sendo voltados para resultados internos, mesmo diante dos

pedidos de estudos sobre a necessidade de considerar uma visão holística em que tanto aspectos internos (*e.g.* características da força de trabalho, rotatividade etc.) quando aspectos externos (*e.g.* percepções do público-alvo e fatores econômicos, sociais e ambientais) fossem considerados (Cooke et al., 2021; Jackson et al., 2014).

Portanto, após trinta anos de produção, teóricos concluem que o tema apresenta tendências controversas, definições incertas, relações mínimas entre GP e desempenho organizacional e pouco rigor metodológico (Cooke et al., 2021; Jackson et al., 2014; Jiang & Li, 2019; Kaufman, 2015a, 2020; Lengnick-Hall *et. al*, 2009; Paauwe & Boselie, 2003, 2005; Paauwe, 2009; Schuler & Jackson, 2014). Alguns autores dão ênfase a essas tendências metodológicas sob as quais o campo de GEP tem se desenvolvido, sobretudo no que tange à “psicologização” dos estudos sobre GP (*e.g.* Budd, 2020; Cascio, 2015; Cooke et al., 2021; Godard, 2014, 2020; Kaufman, 2015a, 2020; Wright et al., 2016).

Nesse sentido, em 2014, Godard publicou uma pesquisa alertando sobre a transição dos estudos sobre GP, os quais, desde 1960, pautavam-se nas relações industriais enquanto base teórica para uma teoria voltada para a psicologia industrial e organizacional, como as teorias de comportamento organizacional. Uma razão central para que essa mudança de perspectiva tenha ocorrido reside no paradigma de pesquisa da utilização da psicologia enquanto “ciência pura” tanto ontologicamente, mas também metodologicamente; isso produz pesquisas estatisticamente rigorosas que atendem a muitos dos critérios para publicação nos principais periódicos da área de negócios e gestão, daí então o argumento de que tais pesquisas passaram a dominar estratégias de publicação (Farndale et al., 2020; Godard, 2014).

A questão é que os problemas de pesquisa relevantes nos campos de GP e das relações de trabalho são muito mais amplos (Farndale et al., 2020). Logo, essa mudança pode representar sérias implicações para a forma com que as organizações enxergam as práticas de GP e os funcionários, uma vez que a unidade de GP pode se tornar massa de manobra em ambientes corporativos que valorizam sistemas de controle e disciplina, tratando funcionários como objeto e suprimindo ainda mais os problemas relacionados às relações de trabalho que envolvem problemas de ordem política (Godard, 2014, 2020).

Em meio ao reconhecimento da influência da Psicologia no campo de estudos de GP, uma edição especial do periódico *Human Resource Management Journal (HRMJ)* publicou quatro artigos que discutem a temática da “psicologização” nos campos de estudos de GP e das relações de trabalho. Um dos estudos é o de Troth e Guest (2020), que rebate as críticas sobre a perspectiva psicológica nas pesquisas sobre GP, principalmente fazendo referência a trechos

do estudo de Godard (2014). Os autores afirmam que a adoção da psicologia pode fornecer contribuições ímpares para produção de estudo multidisciplinar, multinível, multimétodo e *multistakeholders* de GP, uma vez que até aqui, segundo os autores, as demais teorias e campos de estudo adotados pelos estudos sobre GP “falharam em oferecer *insights* relevantes” sobre a relação de GP e desempenho organizacional.

Porém Kaufman (2020), em um desses quatro estudos publicados na mesma edição do periódico *HRMJ*, traça críticas severas ao artigo de Troth e Guest no mesmo ano. Kaufman afirma que, na realidade, o processo de psicologização de GP é responsável por reduzir os resultados macro-organizacionais a aspectos psicológicos dos indivíduos em seu nível mais micro, solicitando que a discussão sobre aspectos de nível macro, voltados à gestão, à estratégia e aos negócios, seja retomada, sem se restringir somente aos aspectos psicológicos.

Nesse sentido, Budd (2020) condensa o embate entre as perspectivas de Kaufman (2015) e Troth e Guest (2019) ressaltando que ambos os interesses são conceitualmente e metodologicamente diferentes, porém isso não deve ser visto como um problema, mas, sim, como uma forma de enriquecer a discussão sobre o tema. Segundo Budd (2020), o problema está em considerar apenas a psicologia, tornando a psicologização uma tendência, podendo se transformar na única abordagem possível para compreender as complexas relações estabelecidas entre as práticas de GP e a organização se largamente adotada.

A partir desse debate, mostra-se evidente que a GEP não pode ser compreendida de forma reducionista, seja considerando apenas o nível macro, voltado às estratégias organizacionais, ou considerando unicamente o nível micro, o qual foca nos aspectos psicológicos individuais. A discussão sobre psicologização das práticas de GP (Kaufman, 2020) e o papel dos sistemas integrados de práticas de GP na promoção de desempenho superior (Lepak et al., 2006; Ostroff & Bowen, 2016) reforça a necessidade de abordar a interligação entre diferentes níveis de análise.

Nesse contexto, a Gestão de Desempenho (GD) emerge como componente central para alinhar práticas de GP com a estratégia organizacional, uma vez que consiste em um sistema estruturante que conecta subsistemas de GP, como treinamento, recompensas e avaliação de desempenho, para maximizar resultados organizacionais (Den Hartog et al., 2004). Dessa forma, a relação entre GP e desempenho mostra-se evidente quando se aplica o sistema de GD, mas também enfrenta muitos dos desafios apontados na literatura sobre GEP e desempenho organizacional.

Portanto, na próxima subseção, serão discutidos os fundamentos do sistema de GD e a sua mensuração, destacando os desafios delineados para a efetividade do sistema nesta revisão e as interconexões com a literatura de avaliação de políticas e práticas de GP.

2.2 Gestão de Desempenho

Como parte da visão da GEP de que gerenciar estrategicamente pessoas é essencial para o desenvolvimento e a sobrevivência das organizações, o processo de avaliar e gerenciar ativamente o desempenho organizacional e do indivíduo passa a fazer parte dos estudos, pois impõe-se que, assim, seria possível promover efetividade organizacional (Brown et al., 2019; Den Hartog et al., 2004).

Nesse sentido, inicialmente, os estudos se referenciam a “gerenciar” o desempenho como sendo a avaliação de resultados, avaliação para pagamento por desempenho, entre outros que faziam menção principalmente à etapa de avaliação, cujas pesquisas datam desde antes de 1920, conferindo pouca ênfase ao gerenciamento em si do desempenho (Den Hartog et al., 2004; DeNisi & Pritchard, 2006).

Avaliar o desempenho é tido como de interesse de profissionais e acadêmicos há mais de 100 anos; porém, mais recentemente, o foco tem se voltado especificamente para gerenciar o desempenho, pois, por mais que os tópicos pareçam semelhantes ou de formas intercambiáveis na literatura acadêmica e na prática, a Avaliação de Desempenho (AD) e Gestão de Desempenho (GD) não são idênticas (DeNisi & Murphy, 2017). A AD é apontada como sendo um evento formal e, usualmente, anual em que a avaliação do desempenho dos funcionários é feita por um juiz, normalmente o seu próprio supervisor, em que não é considerado o alinhamento do desempenho com os negócios e que não fornece feedbacks contínuos capazes de promover desempenho superior (*e.g.* Aguinis, 2013; Schleicher et al., 2018; DeNisi & Murphy, 2017).

Diante disso, com as mudanças no panorama sobre GD principalmente nos últimos dez anos, os estudos têm apontado que a AD é, na realidade, uma parte da GD, pois, como destacado anteriormente, enquanto a AD consiste, normalmente, em um evento pontual, a GD é considerada um processo contínuo composto por diversas atividades com objetivo de gerenciar o desempenho (DeNisi & Murphy, 2017; Kinicki et al., 2013; Levy et al., 2017; Schleicher et al., 2018).

Complementarmente, destaca-se que o sistema de GD é descrito como sendo um “processo contínuo de identificação, mensuração e desenvolvimento do desempenho de indivíduos e equipes alinhado aos objetivos estratégicos da organização” (Aguinis, 2013, p. 2). Essas etapas são descritas na Figura 1 (Armstrong, 2015) sobre o ciclo de GD.

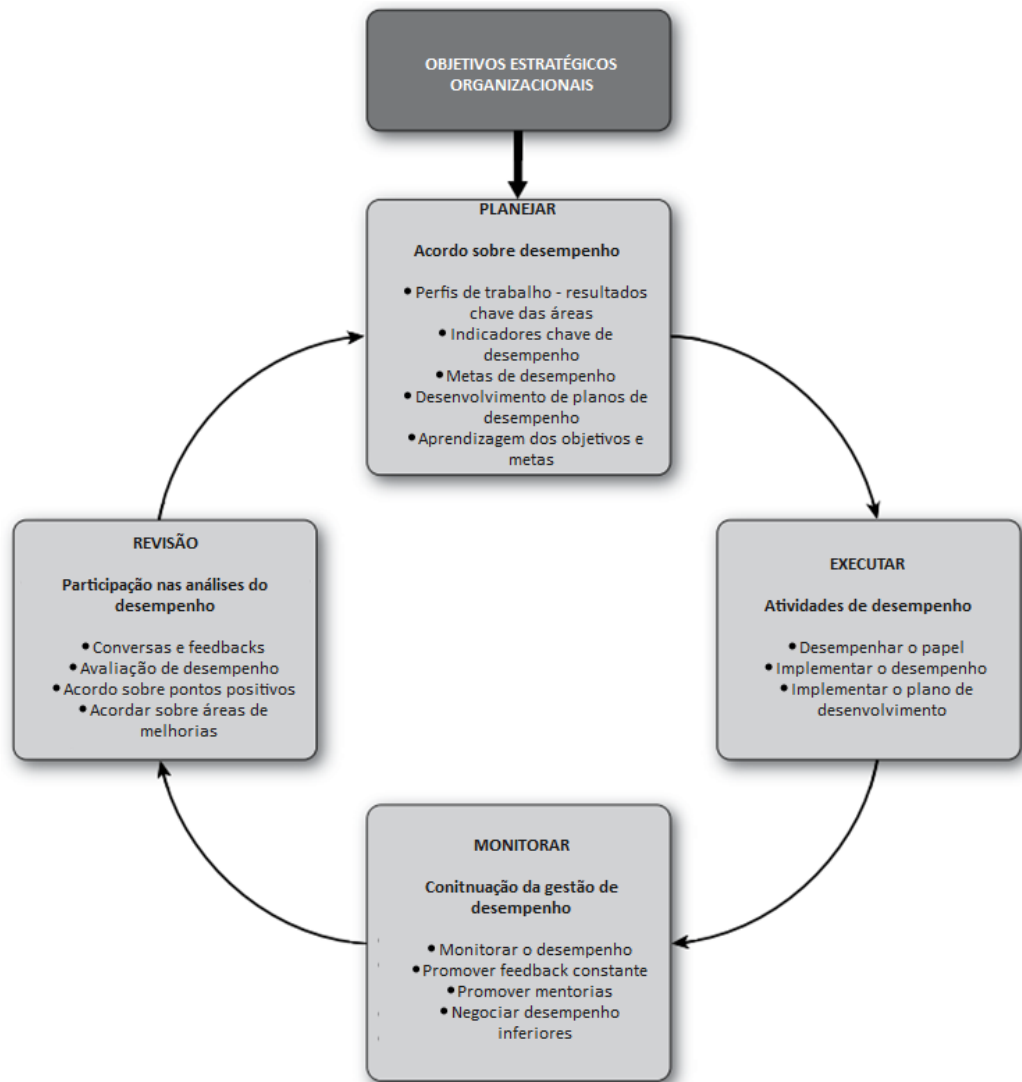


Figura 1. O Ciclo de Gestão de Desempenho

Fonte: Adaptado de Armonstrong (2013, p. 17).

O sistema parte dos objetivos estratégicos organizacionais e segue para quatro etapas de GD em um fluxo contínuo e autorrenovável (Armstrong, 2015). As etapas consistem em: i) planejar, chamado pelo autor de “acordo sobre o desempenho”, em que são formuladas as metas e acordado o desempenho esperado; ii) executar/atuar, em que as demandas de trabalho serão

realizadas, conforme definido no estágio de planejamento, na forma de resultados, metas, requisitos de competência e planos de ação; iii) monitorar, relativo ao processo contínuo de gerenciar o desempenho, onde inclui-se “monitorar regularmente os resultados em relação aos planos e garantir que ações corretivas sejam tomadas quando necessário, realizando monitoramento dos indivíduos e gerenciando o desempenho a partir de feedbacks fornecidos, apoio e orientação”; e iv) revisar, momento realizado uma ou duas vezes ao ano em que as questões-chave de desempenho e desenvolvimento são consideradas e é feita a conclusão do ciclo de GD, fornecendo a base para a atualização dos desempenhos que serão novamente acordados no ciclo seguinte (Armstrong, 2015).

Dessa forma, a GD busca atender a diversos propósitos, sendo que o primeiro deles é apontado por ser auxiliar à alta administração no alcance de objetivos organizacionais por meio do alinhamento com o desempenho individual, seguido de fornecer insumos para tomada de decisão relacionado ao pagamento por desempenho (*e.g.* gratificação financeira e promoção), treinamento e desenvolvimento (*e.g.* capacitação e mentoria) e planejamento da força de trabalho (*e.g.* demissão, contratação e movimentação de pessoal) (Aguinis, 2013).

Para que esses propósitos sejam atingidos, alguns autores discorrem sobre a necessidade de que o modelo da GD atue como um sistema que seja abrangente o suficiente para incorporar todas as práticas de GP necessárias para garantir os meios, a motivação e a oportunidade para promover desempenho organizacional (*e.g.* DeNisi; Smith, 2014); entretanto, outros autores (*e.g.* Aguinis, 2013) entendem de forma distinta e alertam para a necessidade de impor limites sobre o que está sendo considerado em um sistema de GD ao não se excluir praticamente nenhuma prática de GP do modelo de sistema de GD (Schleicher et al., 2018).

Percebe-se que, apesar da divergência sobre o que deve ser considerado em um modelo de GD, os autores concordam sobre a importância de ele ter sido como um sistema em que seus elementos – muitas vezes subsistemas de GP – são interrelacionados de forma que uma mudança em um componente impactará no outro. Segundo Schleicher et al. (2018), estudar a GD enquanto um sistema permite ampliar conhecimento sobre a efetividade de GD, pois fornece insumos que auxiliam a compreender relações com múltiplos componentes e, por isso, também suas relações.

Dessa forma, estudos indicam que grande parte das organizações utilizam o sistema de GD para mensurar o desempenho individual, promovendo feedbacks contínuos e políticas de recompensa, promoção ou desenvolvimento (*e.g.* Mercer, 2013; Cascio, 2006; Levy & William, 2004). Destaca-se que, para além das organizações privadas, organizações públicas de diversos

governos ao redor do globo têm adotado a GD como parte da GP (e.g. Abu-Shanab et al., 2021; Hondegheem & Dorpe, 2013; Longenecker & Nykodym, 1996; Sotirakou & Zeppou, 2006).

No Brasil, a GD também é adotada nos órgãos da Administração Pública face ao contexto de reformas gerenciais empreendidas a partir do final do século XX (Coelho et al., 2016; Martins & Ensslin, 2020; Odelius & Santos, 2007). Assim, tem-se uma série de mudanças a partir da adoção de um modelo pós-burocrático de gestão pública que se baseia em valores profissionalizantes e de competitividade decorrentes do movimento da Nova Administração Pública (ou *New Public Management* – NPM), que prevê a adoção no setor público de políticas, práticas e valores gerenciais típicos do setor privado (Barzelay & Jacobsen, 2009; Hood, 1995; Secchi, 2009).

No caso do setor público brasileiro, o sistema de GD também reflete o seu histórico de desenvolvimento que se inicia com a inserção da prática de AD, a qual foi prevista em 1998 na Emenda Constitucional nº 19 (Martins et al., 2010). Em complementação, após dez anos, foi promulgada a Lei nº 11.784/08, a qual define a avaliação de desempenho, para os fins previstos na Lei, a avaliação de desempenho, como sendo “o monitoramento sistemático e contínuo da atuação individual do servidor e institucional dos órgãos e das entidades, tendo como referência as metas globais e intermediárias dos órgãos e entidades que compõem o Sistema de Pessoal Civil” (Brasil, 2008).

As diretrizes sobre a AD, conforme disposto na Lei nº 11.784/08, somente foram regulamentadas em 2010 a partir da edição do Decreto nº 7.133/10, em que foram dispostos os critérios e procedimentos gerais a serem observados pelas unidades responsáveis pela AD dos órgãos públicos. Dessa forma, ampliou-se o espaço para que a GD na Administração Pública fosse instituída nas organizações, tendo em vista que o Decreto conferiu maior espaço de atuação ao alternar da AD para a GD, não restringido mais somente à etapa de avaliação, mas incluindo as etapas de planejamento e monitoramento. Para complementar, foi publicado, em 2013, o Manual de Orientação para a Gestão de Desempenho com as orientações detalhadas sobre o processo de GD no setor público (Brasil, 2010).

Após quase dez anos desde a edição do Manual, o tema tornou a fazer parte das discussões no setor público brasileiro com a Proposta de Emenda à Constituição – PEC nº 32, de 2020 (Brasil, 2020), que insere um sistema de GD definitivo com formulação de metas individuais e setoriais, capaz de regulamentar o que já possui viabilidade legal e tornaria possível realizar o desligamento de servidores públicos em razão de desempenho individual

insuficiente (Secchi et al., 2021; Vilhena & Martins, 2022). A proposta segue em discussão, mas, se aprovada, pode conferir ainda mais peso ao sistema de GD na Administração Pública.

Apesar de previsão legal para a implementação da GD nas organizações públicas, destaca-se que os desafios para uma implementação efetiva ainda são muitos, sobretudo em razão de interferentes contextuais próprios do setor público que interferem na implementação, como também das dificuldades na definição de objetivos e parâmetros avaliativos adequados à realidade das organizações públicas (Araújo, 2018; Martins et al., 2009).

Logo, é interessante destacar que a GD tem sido adotada formalmente pela maioria das grandes organizações, sejam elas públicas ou privadas, e continuam atuais e recorrentes. Isso acontece mesmo em um cenário de constante reclamação e insatisfação em relação à GD por parte dos funcionários (Murphy, 2020). As pesquisas sobre GD ao longo do tempo mostram que, apesar do aprofundamento teórico que inclui a distinção entre os conceitos de AD e GD, buscando delimitar a relevância e o papel específico de cada uma, na prática, tanto funcionários quanto gestores frequentemente percebem ambas como insuficientes ou problemáticas (Murphy, 2020; Schleicher et al., 2018).

Levy et al. (2017) afirmam que gerenciar o desempenho individual não tem funcionado nas organizações e, por mais que se saiba disso há certo tempo, somente agora isso tem preocupado praticantes e acadêmicos para uma mudança de postura, a qual não envolve apenas mudanças operacionais ou metodológicas na forma com que as avaliações de desempenho são conduzidas. Dessa forma, é comum que acadêmicos e praticantes questionem se todo investimento feito para implementar e promover a GD nas organizações realmente traz benefícios substanciais à organização (Murphy, 2020).

O sistema de GD permanece sendo uma prática cujos resultados se mostram distantes sob o olhar de atores organizacionais (Cappelli & Conyon, 2017). Nesse sentido, passou-se a questionar os reais efeitos daquilo que era prescrito pela literatura de GD (Lawler & McDermott, 2003) e, apesar de um progresso significativo ter sido feito em estabelecer a relação entre GP e desempenho, problemas teóricos e empíricos permanecem (Den Hartog et al., 2004). Mesmo com a onipresença de GD na academia e na prática, muitos sistemas de GD mostram-se operados distantes da perfeição, restando uma série de questões não respondidas acerca da efetividade de GD (Armstrong, 2015; Ikramullah et al., 2016; Schleicher et al., 2018).

2.2.1 Efetividade de Gestão de Desempenho

A efetividade do sistemas de GD pode ser definida como o alinhamento entre funcionários e objetivos organizacionais (Armstrong, 2015). Entretanto, isso não é tão simples assim: demonstrar a efetividade, em específico, de sistemas de GP mostra-se uma tarefa complexa e de difícil execução, mais até do que a acessar a efetividade de outras práticas organizacionais que não são de GP (Ahmed, 1999).

Passam, então, a vigorar questões quanto à capacidade do sistema de GD em promover efetividade organizacional (*e.g.* Awan et al., 2020; Cappelli & Conyon, 2017; Dewettinck, 2008; Dewettinck & van Dijk, 2013; Iqbal et al., 2015, 2019; Kakkar et al., 2020; Keeping & Levy, 2000; Lawler, 2003; Modipane et al., 2019a; Murphy & Cleveland, 1995; Pichler, 2012; Schleicher et al., 2018, 2019; Sharma et al., 2016; Tuytens & Devos, 2012). Essas questões residem, em grande parte, nos mesmos problemas de se avaliar a efetividade políticas ou práticas de GP (Den Hartog et al., 2004).

Nesse sentido, pesquisas têm buscado investigar a efetividade dos sistemas de GD utilizando certos critérios de mensuração e de resultado para acessar as relações antecedentes- resultados que manifestam a efetividade de GD (Iqbal et al., 2015). Para isso, a atenção ao longo do tempo tem se deslocado para diferentes critérios de avaliação – erros de avaliação, utilização, contexto social, entre outros (Cawley et al., 1998; Keeping & Levy, 2000). Inclui-se, ainda, a recorrência com que os estudos aplicam os mesmos critérios, como no caso de indicadores financeiros ou a percepção sobre o sistema de GD, representada pela dimensão de “reação”, resultando em um foco desproporcional por determinados critérios adotados (Schleicher et al., 2019).

Críticas mais recentes incidem na falta de interação entre esses critérios, que são, muitas vezes, adotados de forma isolada e distantes de uma abordagem holística (Ikramullah et al., 2016; Iqbal et al., 2015, 2019; Levy & Williams, 2004; Schleicher et al., 2019). Os estudos que apontam essa lacuna são os mesmos que buscaram preenchê-la ao estabelecer modelos integrativos de avaliação de efetividade de GD ao complementarem o campo com outras visões teóricas.

Esse é o caso do modelo de avaliação de efetividade de avaliação de desempenho proposto por Levy e Williams (2004), conforme a Figura 2. O estudo feito pelos autores realiza uma revisão de literatura para levantar aspectos relacionados à teoria de contexto social (*Social Context Theory – SET*), que discorre sobre a influência do contexto social, composto por

diferentes atores organizacionais, em práticas e políticas organizacionais. Assim, segundo os autores, o contexto social em que a AD está inserida desempenha um papel importante na efetividade do processo de avaliação, o que influencia como os participantes da avaliação reagem a esse processo. O modelo de avaliação de efetividade proposto pelos autores se pauta na reação dos participantes da AD, corroborando a produção crescente sobre o tema à época em que as reações assumem um importante papel enquanto critérios de avaliação de efetividade. Além disso, os autores entendem que essa é a forma com que os participantes mais conseguem enxergar os benefícios da AD. Os autores também consideraram erros e vieses do avaliador e a precisão da avaliação enquanto critérios de avaliação de efetividade de AD.

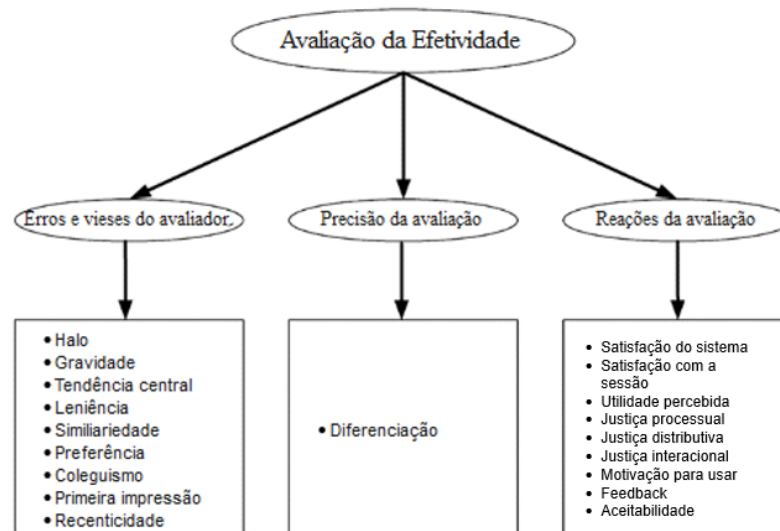


Figura 2. Avaliação de efetividade

Fonte: Traduzido de Levy e Williams (2004).

Sob certa influência do modelo anterior, Iqbal et al. (2015) propõem um modelo integrado que também se baseia nas reações de avaliados para avaliação de efetividade da AD. O modelo foi desenvolvido a partir do enfoque dado por alguns estudos (*e.g.* Keeping & Levy, 2000; Levy & Williams, 2004; Pichler, 2012), em que o centro da análise são as reações dos avaliados. O modelo proposto é composto de quatro partes (propósito, precisão e justiça da AD e suas subpartes) que consideram diferentes aspectos. Assim, destaca-se que as três primeiras partes são referentes aos critérios de avaliação que se relacionam às reações dos avaliados, e a

quarta e última parte (central) aborda a correlação entre os critérios de avaliação de efetividade de AD (critérios relacionados às reações sobre a organização e sobre o indivíduo). Segundo os autores, a interpretação utilizada para desenvolver o modelo promove evidência empírica e teórico-racional de que os critérios de avaliação de efetividade estão relacionados à reação dos avaliados.

O modelo proposto para avaliação de efetividade da AD dos autores Ikramullah et al. (2016) utiliza a Teoria de Valores Concorrentes (*Competing Values Framework* – CVF). O CVF foi desenvolvido por pesquisadores organizacionais com intuito de demonstrar fatores que influenciam na efetividade organizacional em que são dispostas quatro dimensões de valor competitivo em quadrantes opostos: interno vs externo; controle vs flexibilidade; meio vs finalístico (Ikramullah et al., 2016). Assim, inspirados pelo CVF, os autores propõem processo interno, relações humanas, sistema aberto e modelo racional de metas como importantes aspectos para se avaliar a efetividade de AD/GD.

O modelo mais recente foi proposto por Schleicher et al. em 2019 a partir de uma revisão teórico-conceitual dos últimos 30 anos de produção científica sobre AD/GD, incluindo tanto pesquisas empíricas como conceituais com viés acadêmico e prático sob a lente micro e macro, de forma a obter um panorama sobre os critérios utilizados para mensurar e discutir a efetividade de AD/GD. Assim, os autores propõem um modelo que precisa ser “explicitamente abrangente” a partir de uma abordagem interativa em que se levou em consideração três aspectos: a) o conteúdo existente no campo de avaliação; b) critérios de interesse de praticantes e acadêmico; e c) construtos de nível micro e macro-organizacional.

Atendendo aos anseios descritos nos três aspectos apontados anteriormente, os autores desenvolveram um modelo de avaliação de efetividade de GD, contido na Figura 3. Para isso, foram incorporadas teorias de avaliação tanto da própria literatura de GD (*e.g.* Cardy; Dobbins, 1994; Keeping & Levy, 2000; Levy & Williams, 2004) como também de GP, em especial na literatura de avaliação de treinamento (*e.g.* taxonomia de Kirkpatrick) para a construção do modelo dedutivo-integrativo; outro ponto atendido pelo modelo foi o de incorporar aspectos que tanto eram relacionados à academia como à prática em razão do *gap* existente entre estes dois campos, tendo em vista a diferença de foco existente com relação aos critérios de resultados; e, por último, o modelo proposto pelos autores buscou ser abrangente e generalista a fim de incorporar critérios provenientes de níveis de análise de distintos critérios, como “micro” e “macro” organizacional, para que dessa forma seja possível examinar de que forma a GD influencia resultados individuais e organizacionais.

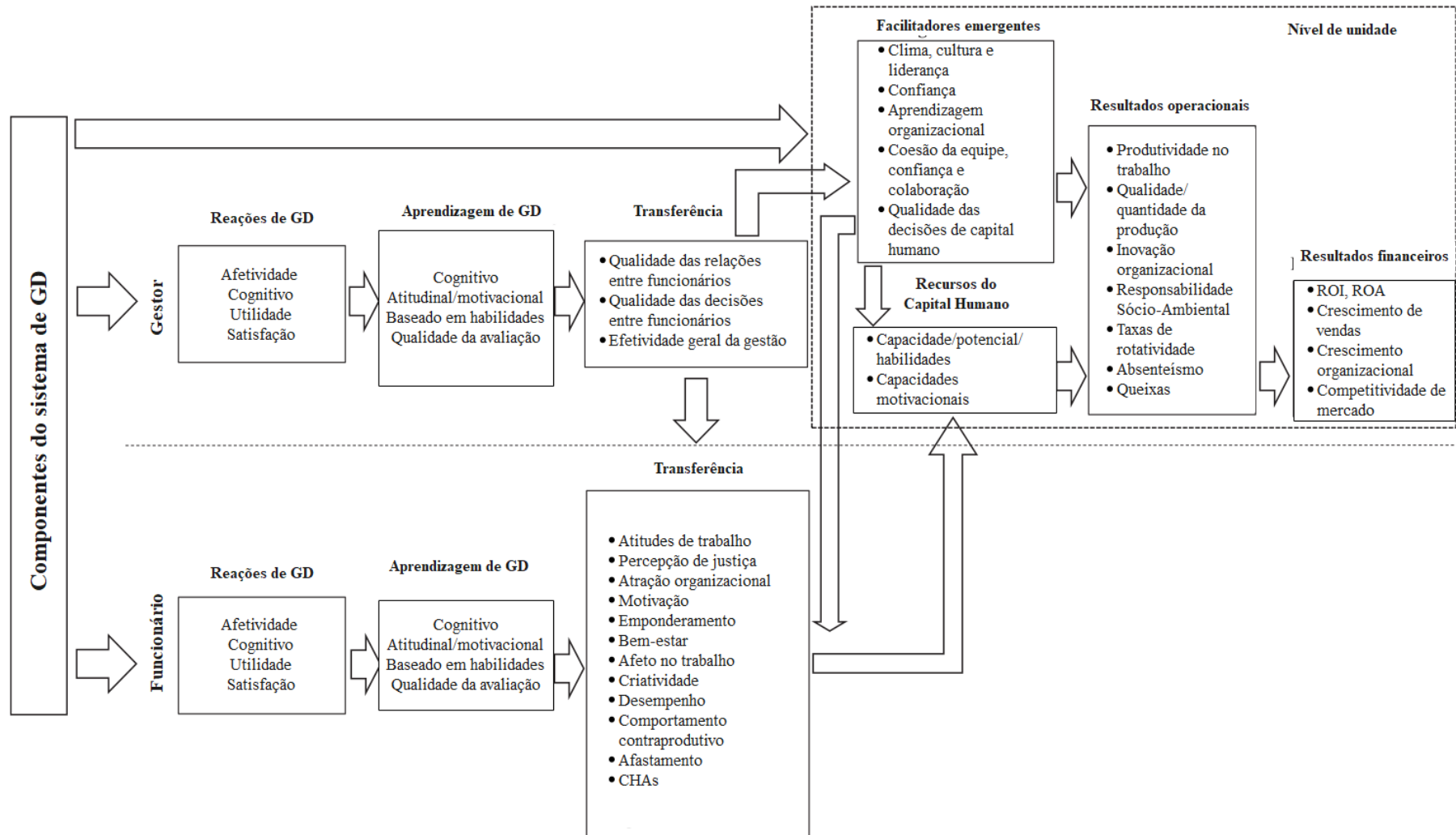


Figura 3. Modelo de critérios avaliativos da efetividade de Gestão de Desempenho

Fonte: Adaptado de Schleicher et al. (2019, p. 854).

Portanto, conclui-se que, apesar desses modelos, o que há é um acúmulo de estudos acerca da avaliação de efetividade de GD que não se integram, o que prejudica o desenvolvimento do campo científico e prático em razão da ausência de teorias e métodos objetivos e evidentes, cujos critérios realmente forneçam contribuições na identificação da efetividade do sistema de GD. Mais do que isso: a falta de uma sistematização dos estudos acerca dos critérios de efetividade de GD dificulta o conhecimento acumulativo do campo sobre os critérios relevantes e como eles podem contribuir para a identificação dos resultados relacionados ao desempenho organizacional.

A sistematização da produção torna possível refletir acerca das possibilidades de critérios para avaliação de efetividade de GD, delineando aqueles mais consolidados e os ainda marginalizados pela literatura de GD, mas necessários para compreensão da efetividade de GD. Com esses insumos, espera-se que tanto pesquisadores quanto profissionais possam, mesmo reconhecendo as fragilidades do sistema de GD, identificar quais resultados organizacionais o sistema consegue entregar e como.

Portanto, empreendeu-se por uma revisão sistemática da literatura dos modelos, dimensões e critérios adotados nos estudos dedicados à avaliação de efetividade dos sistemas de GD. A revisão sistemática será apresentada na subseção seguinte.

2.3 Proposição de um Modelo Teórico-Integrativo de Avaliação da efetividade do sistema de Gestão de Desempenho

Para construção do modelo teórico-integrativo de avaliação da efetividade do sistema de Gestão de Desempenho, inicialmente empreendeu-se uma revisão sistemática de literatura dos critérios utilizados para avaliação de efetividade dos sistemas de GD nas bases de dados: *Scopus* e *Web of Science* – com resultados apenas para os descritores em inglês – e *Scielo*, *Spell*, *Periódicos Capes* e *Google Acadêmico* – para os resultados para os termos em português.

A busca foi feita, sem restrição de início, até julho de 2024, considerando o título, resumo e palavras-chave com os seguintes descritores: “*performance management*”, “*performance appraisal*” e “*performance evaluation*” combinados com o termo “*effectiveness*”; e “gestão de desempenho” e “avaliação de desempenho” em conjunto com a palavra “efetividade”. Consideraram-se os termos AD e GD de forma intercambiável; apesar da diferença conceitual existente e indicada no início deste trabalho, considerou-se por bem utilizar

ambos os termos nas buscas, evitando a exclusão de estudos que porventura adotassem os termos como semelhantes.

Inicialmente, foram obtidos mais de 18 mil resultados durante a aplicação da combinação desses termos nas bases de dados, porém percebeu-se que a maior parte dos resultados se distanciava dos termos desejados de pesquisa. Para que se pudesse identificar de forma mais eficiente os artigos que se aproximavam dos termos pesquisados, foi feita a ordenação dos resultados por relevância ainda nas bases de dados.

Fez-se então a leitura dos títulos, resumo e palavras-chave dos artigos resultantes das buscas em cada base de dados até que não restasse mais correspondência com os termos pesquisados. Assim foi feita a leitura de quase 2.000 registros, resultando na seleção de 271 artigos internacionais.

Em seguida, para garantir transparência e otimizar o processo de seleção dos artigos da revisão (Page et al., 2021), foi aplicado o guia de *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA), conforme processo indicado na Figura 4.

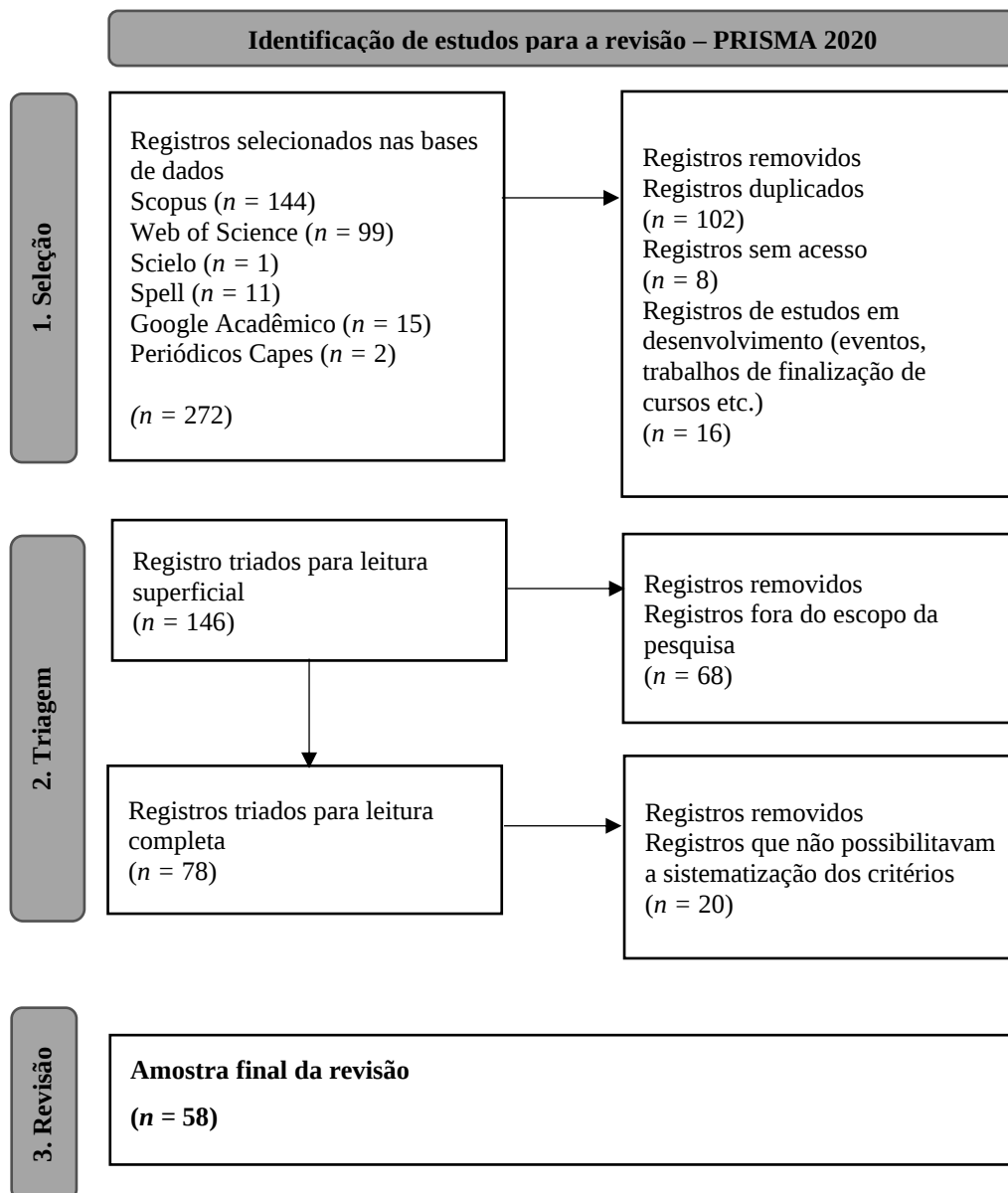


Figura 4. Diagrama utilizado para revisão sistemática de literatura

Fonte: Adaptado de *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA) (2020).

O primeiro filtro aplicado nos artigos identificados foi a exclusão dos estudos duplicados, inacessíveis e em desenvolvimento. Nesse processo, resultaram 146 registros de estudos que foram lidos de forma superficial, apenas para confirmar a relação com objeto de estudo da pesquisa, excluindo 68 artigos. Com a amostra de 78 artigos, foi feita uma nova leitura dos textos, desta vez integralmente, seguido pelo fichamento dos estudos; entretanto, em 20 artigos não foi possível a identificação de critérios ou outras formas de avaliação de efetividade

de GD, e esses artigos foram excluídos da amostra. Por fim, a amostra da revisão foi composta por 58 artigos.

Com a amostra da revisão selecionada, foi feita a identificação de todos os critérios adotados em cada um dos estudos revisados para avaliação de efetividade dos sistemas de GD. Inicialmente, a identificação foi feita a partir da própria declaração feita pelos autores ao longo do texto do estudo revisado, porém, quando não foi possível o reconhecimento em razão da falta de clareza ou da ausência dessa informação, a autora desta pesquisa analisou aspectos metodológicos do estudo, como o instrumento de pesquisa aplicado.

É importante pontuar que há, em alguns estudos revisados, uma certa confusão teórica entre o que é a variável ou medida independente “efetividade de GD” (*e.g.* percepção de efetividade do sistema de GD, em Kakkar et al. (2020); Questionário de Efetividade do Sistema de Gestão de Desempenho, em Mahbubela et al., 2016), cuja operacionalização envolvia, em alguns casos, mais de um critério; e o que é um modelo de avaliação de efetividade de GD (*e.g.* Levy & Williams, 2004; Schleicher et al., 2019), normalmente com uma visão mais holística da efetividade em que foram considerados diferentes aspectos, incluindo nível de análise micro e macro-organizacional, resultados mais próximos ou distantes do sistema de GD e considerando ambiente interno e externo à organização. Em alguns casos, esses aspectos se misturam, o que tornou a identificação das dimensões e dos critérios de avaliação dificultosa em certos momentos da codificação da literatura revisada.

Diante disso, todos os critérios adotados pelos estudos foram registrados em uma planilha eletrônica em que se iniciou a classificação destes por meio de análise de conteúdo pós-categórica (Bardin, 2016). Na medida em que as categorias emergiam dos estudos a partir dos critérios identificados, fez-se necessário aglutinar, suprimir ou desmembrar algumas categorias inicialmente propostas de forma a atender às regras de categorização de Bardin (2016). Deve-se destacar que todos os critérios identificados nos estudos revisados foram considerados em ao menos uma categoria.

As categorias seguiram os princípios definidos por Bardin (2016) de i) *exclusão mútua*, não havendo a representação de um mesmo elemento em diferentes categorias propostas; ii) *homogeneidade*, em que a mesma categoria deve funcionar sob um mesmo registro e nível de análise; iii) *pertinência*, em que uma categoria foi considerada pertinente quando se mostrou representativa em relação ao material de análise obtido, refletindo as intenções desta investigação; iv) *objetividade e fidelidade*, em que diferentes partes de um mesmo material seguiram a mesma codificação, permitindo que outras análises possam ser feitas e o mesmo

resultado deverá ser obtido; e, por último, v) *produtividade*, ressaltando que cada categoria proposta produz resultados férteis para o que está a se investigar.

Antes de descrever os critérios identificados, deve-se pontuar que, para melhor organização dos resultados e seguindo a proposta de modelos de avaliação de resultados de sistemas e práticas de GP apresentados anteriormente (*e.g.* Schleicher et al., 2019; Dyer & Reeves, 1995; Paauwe, 2009), considerou-se por bem agrupar os critérios em dimensões conforme sua natureza. Assim, seis dimensões foram propostas:

- i. *Reação*, referente às impressões acerca do sistema de GD (Iqbal et al., 2015; Keeping & Levy, 2000; Levy & Williams, 2005; Schleicher et al., 2019);
- ii. *Aprendizagem*, referente aos conhecimentos, habilidades e atitudes adquiridos pelo indivíduo a partir da participação no sistema de GD (Schleicher et al., 2019);
- iii. *Transferência*, referente à capacidade em transferir resultados obtidos por meio do sistema de GD para o trabalho (Schleicher et al., 2019);
- iv. *Resultados Operacionais*, referentes à operação de processos organizacionais;
- v. *Resultados Financeiros*, que consistem em orçamento, ações, produtos e serviços financeiros, econômicos ou contábeis (Nankervis et al., 2012); e
- vi. *Impacto Societal*, relativo às contribuições do sistema de GD em relação ao que é entregue pela organização ao seu mundo externo, incluindo *stakeholders* (Ikramullah et al., 2016).

Essas dimensões seguem a ordem lógica de disposição no modelo, a qual utiliza como critério a emergência dos resultados do sistema de GD, partindo dos mais imediatos e próximos ao sistema de GD para aqueles menos, tal como apresentado tanto nos modelos de avaliação dos resultados de GD (*e.g.* Levy & Williams, 2004; Schleicher et al., 2019) quanto por aqueles que buscam mensurar os resultados dos sistemas e das práticas de GP (*e.g.* Dyer & Reeves, 1995; Kaufman, 2020; Paauwe, 2009).

Assim, cada uma das dimensões propostas é composta por critérios identificados na literatura revisada. Os critérios permitem acessar o resultado do sistema de GD, representando a sua efetividade (Iqbal et al., 2015). Com isso, a categorização da produção científica dos 58 artigos revisados foi sistematizada em 22 critérios identificados por terem sido adotados para avaliação de efetividade do sistema de GD. As definições de cada um dos critérios identificados podem ser conferidas na Tabela 1.

Tabela 1

Definições dos critérios identificados para avaliação de efetividade de GD

Dimensão	Critério	Definição operacional
Reação	Justiça	Extensão com que os funcionários percebem o sistema de GD como justo e igualitário (Sharma et al., 2016)
	Precisão	Capacidade de refletir corretamente aquilo a que ele se propõe a fazer (Sharma et al., 2016)
	Satisfação	Nível de contentamento geral com o sistema de GD (Schleicher et al., 2019; Culbertson et al., 2013)
	Utilidade	Capacidade do sistema de GD em fornecer insumos para outras políticas e práticas da organização (Ikramullah et al., 2016)
	Afetividade	Sentimentos experienciados pelos usuários em relação ao sistema de GD (Schleicher et al., 2019)
Aprendizagem	Atitude	Intenção de se comportar ou agir em decorrência da interação com o sistema de GD (Schleicher et al., 2019)
	Habilidade	Capacidade de aplicar os conhecimentos adquiridos com o sistema de GD (Ikramullah et al., 2016; Lawler, 2003; Schleicher et al., 2019)
	Conhecimento	Informações declarativas, procedimentais ou tácitas adquiridas em decorrência da interação com o sistema de GD (Schleicher et al., 2019)
Transferência	Desempenho no trabalho	Comportamento expressos no trabalho voltados para o alcance de resultados organizacionais (Campbell, 1993; Motowilde, Borman & Schmit, 1997)*
	Motivação	Disposição ou a tendência em exercer esforços no trabalho conforme procedimentos e políticas da organização (Jiang et al., 2013; Latham, Locke & Fassina, 2005)*
	Atitudes no trabalho	Sentimentos, crenças e apego ao trabalho (Judge & Kammeyer-Muller, 2012)*
	Relação Interpessoal	Interação social entre os indivíduos da organização (Schleicher et al., 2019)
	Bem-estar	Aumento do prazer e a redução do sofrimento no trabalho (Schleicher et al., 2019)
Resultados Operacionais	Produtividade	Capacidade de realizar uma quantidade maior de trabalho com uma quantidade igual ou menor de recursos (Schleicher et al., 2019)
	Rotatividade	Quantidade de admissões, demissões e absenteísmo do funcionário na organização (Schleicher et al., 2019)
	Inovação	Desenvolvimento de novas ferramentas ou mudanças no processo de trabalho decorrentes do sistema de GD (Schleicher et al., 2019)

	Força de Trabalho	Qualidade e/ou quantidade do quadro de funcionários da organização (Nankervis et al., 2012; Schleicher et al., 2019; Stanton & Nankervis, 2011)
Resultados Financeiros	Lucratividade	Indicador do ganho financeiro obtido, retirando-se tributos, custos e despesas (Nankervis et al., 2012; Stanton; Nankervis, 2011)
	Retorno Financeiro	Indicador do ganho financeiro obtido após um investimento (Nankervis et al., 2012; Schleicher et al., 2019; Stanton; Nankervis, 2011)
	Participação no Mercado	Parcela de determinado produto ou serviço que a organização detém no mercado (Nankervis et al., 2012; Stanton; Nankervis, 2011)
Impacto Societal	Satisfação do público-alvo	Nível de contentamento do público-alvo com a organização (Ikramullah et al., 2016)

Nota: As definições com * são referentes ao estudo que inicialmente as cunhou e que foi citado por estudo revisado.

As dimensões e seus respectivos critérios foram dispostos em um modelo teórico-integrativo de avaliação de efetividade do sistema de Gestão de Desempenho proposto neste estudo, cujo desenho está na Figura 5. Além das dimensões e critérios, o modelo representa como esses aspectos ocorrem em três níveis de fenômeno: individual, organizacional e ambiental. Nesse sentido, o modelo parte do nível micro, referente ao indivíduo, cujos resultados são mais imediatos e proximais ao sistema de GD, segue em direção às dimensões de nível macro, referente à unidade/organização, e finaliza no nível ambiental/societal, em que os resultados são os mais indiretos e distais do sistema de GD, seguindo os modelos de avaliação de resultados multinível de GEP (*e.g.* Banks; Kepes, 2015; Jiang & Li, 2019; Jiang et al., 2013), apontados ao longo do capítulo anterior.

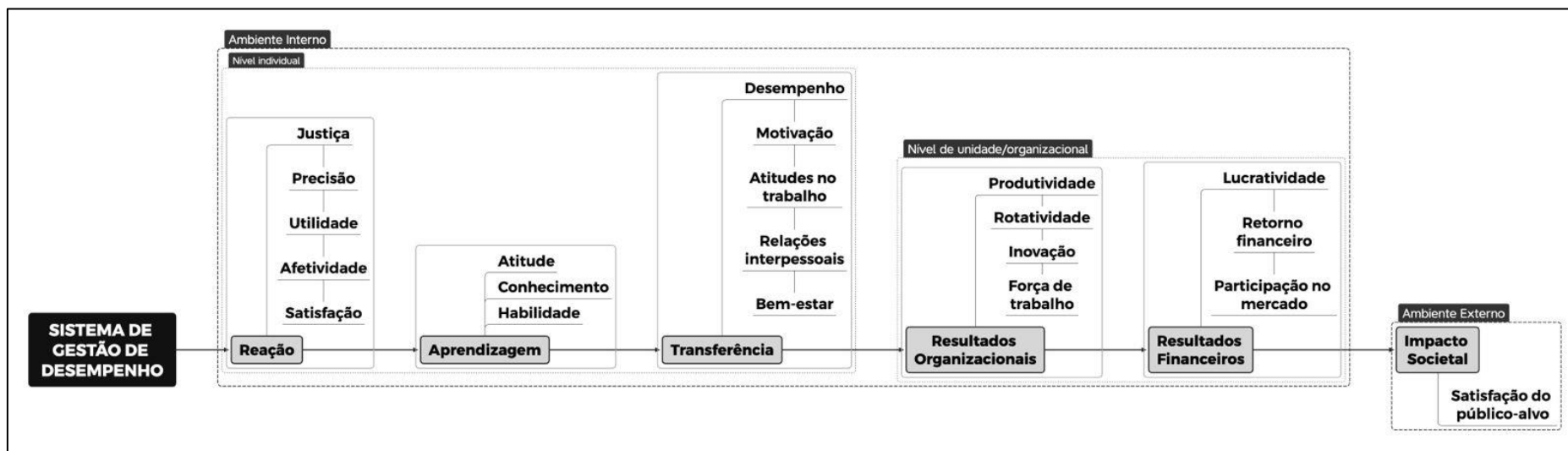


Figura 5. Modelo teórico-integrativo de avaliação de efetividade do sistema de GD.

Fonte: Elaborado pela autora.

Destaca-se como limitação desse modelo a ausência das variáveis contextuais ou de insumo, indicadas em alguns modelos de avaliação de efetividade de GD (*e.g.* Schleicher et al., 2019; Ikramullah et al., 2016), que consideram aspectos como tecnologia, estrutura organizacional, cultura, reputação das unidades de GP etc., os quais são apontados como interferentes no processo de GD (cf. Araújo, 2018). Porém esse modelo teórico-integrativo considerou dimensões e critérios para acessar a relação antecedente-resultado que manifesta a efetividade de GD, conforme apontado por Iqbal et al. (2015).

A seguir, serão exploradas as dimensões e os critérios do modelo teórico-integrativo proposto, com o objetivo de sistematizar as diversas medidas disponíveis na literatura. Essa sistematização busca reduzir a dispersão conceitual e, ao mesmo tempo, identificar quais dimensões e critérios podem contribuir para uma análise mais robusta da relação entre os resultados do sistema de GD e o desempenho organizacional. Dessa forma, pretende-se definir os elementos mais relevantes a serem utilizados na avaliação da efetividade da GD, com base em contribuições teóricas e empíricas presentes na literatura.

2.3.1 Medidas do Modelo Teórico-Integrativo de Avaliação de Efetividade do Sistema de Gestão de Desempenho

A Tabela 2 contempla a frequência com que cada dimensão e seus respectivos critérios foram adotados. Identificou-se que um mesmo estudo, por vezes, adotou mais de uma dimensão ou critério, pois estes não foram utilizados de forma exclusiva pelas pesquisas, mas, sim, complementarmente e em diferentes arranjos. Logo, um mesmo estudo pode figurar em mais de uma dimensão ou de um critério.

Tabela 2

Dimensões e critérios de avaliação de efetividade de GD

Nível do fenômeno	Dimensão	Frequência		Critério	Frequência	
		<i>n</i>	%		<i>n</i>	%
Individual	Reação	41	72	Justiça	24	42
				Precisão	20	35
				Satisfação	14	25
				Utilidade	13	23
				Efetividade	8	14
	Aprendizagem	31	54	Atitude	18	32

Unidade/ Organizacional	Transferência	40	68	Habilidade	17	30
				Conhecimento	13	23
				Desempenho no trabalho	21	36
				Motivação	18	32
				Atitudes no trabalho	16	28
				Relação Interpessoal	11	19
				Bem-estar	3	5
	Resultados Operacionais	16	28	Produtividade	10	18
				Rotatividade	7	12
				Inovação	6	11
				Força de Trabalho	4	7
	Resultados Financeiros	3	5	Lucratividade	3	5
				Retorno Financeiro	3	5
				Participação no Mercado	3	5
Ambiental	Impacto Societal	4	7	Satisfação do público-alvo	4	7

A sistematização dos estudos revisados demonstrou que a dimensão Reação, adotada em 41 artigos, é a mais utilizada para se realizar uma avaliação de efetividade do sistema de GD. De forma semelhante, o critério mais adotado pelos estudos foi Justiça, adotado em 24 artigos, o qual pertence à dimensão Reação. Além dessa dimensão, identificou-se as dimensões Transferência ($n = 40$), Aprendizagem ($n = 31$), Resultados Operacionais ($n = 16$), Resultados Financeiros ($n = 3$) e Impacto ($n = 4$), respectivamente.

Diante dos resultados, nota-se uma tendência do campo em avaliar a efetividade do sistema de GD a partir de uma predominância por critérios específicos, convergindo com o que foi apontado por Ikramullah et al. (2016) ao afirmarem que as pesquisas têm voltado sua atenção de forma desigual para critérios específicos, o que, segundo os autores, pode trazer consequências para a pesquisa e os praticantes ao deslocar a atenção somente para os mesmos critérios. Tal restrição a apenas alguns critérios para acessar os benefícios e as implicações de uma prática de GP em diferentes tipos de organizações e, em diferentes contextos organizacionais, pode mascarar a importância distinta existente entre as organizações (Colakoglu et al., 2006).

Além disso, destaca-se a predominância dos estudos por dimensões cujos fenômenos são de nível individual, como Reação, Transferência e Aprendizagem, e, por consequência, do lado oposto, a perspectiva organizacional, como resultados operacionais e financeiros, é adotada em menor proporção do que os individuais. Esses resultados reforçam o que foi

apontado anteriormente por Budd (2020), Godard (2014) e Kaufman (2020) em relação à psicologização do campo de GP, em que os estudos têm recorrentemente focado, de forma exclusiva ou principal, na perspectiva interna e micro-organizacional para demonstrar os resultados de GP, mesmo quando existem outras abordagens não psicológicas passíveis de mensuração.

Destaca-se que as dimensões e os critérios foram adotados de forma combinada, ou seja, a maioria (74%) dos estudos revisados utilizou mais de uma dimensão ou critério para avaliar a efetividade de GD. A adoção conjunta de diferentes dimensões consiste em um dos principais achados da revisão realizada, pois enriquece a discussão sobre a necessidade de que diferentes medidas sejam consideradas para avaliar a efetividade de GD. Esse resultado corrobora a literatura produzida até então em que os autores assumem não ser possível definir um “melhor critério” para mensurar a efetividade em razão do seu caráter multidimensional (Ahmed, 1999; Connolly et al., 1980; Ikramullah et al., 2016; Iqbal et al., 2015, 2019; Keeping & Levy, 2000; Lawler, 2003; Nagel, 1986; Schleicher et al., 2019). Portanto, realizar um estudo com intuito de avaliar a efetividade de GD a partir de apenas uma dimensão ou um critério significa desconsiderar os demais resultados do sistema de GD que compõem a efetividade.

Essa multidimensionalidade do construto “efetividade” também dá luz a como cada resultado adotado para avaliar o sistema de GD pode variar a depender das necessidades, valores e preferências de grupos de *stakeholders* da organização (executor das atividades, gerente, diretor, acionista, consumidor, cidadão, sindicalista etc.), o que reforça a complexidade de determinar a efetividade de GD (Colakoglu et al., 2006; Ikramullah et al., 2016).

A partir da revisão sistemática de literatura empreendida, pôde-se também identificar o comportamento metodológico dos estudos na consecução de seus objetivos em avaliar a efetividade do sistema de GD. Dessa forma, foi possível identificar as medidas utilizadas para avaliar cada critério identificado nos estudos revisados, assim como a indicação do nível da medida adotada, tendo sido coletada no nível do indivíduo (individual) ou um dado de nível de unidade/organizacional. Também foi analisada a natureza da medida utilizada: subjetiva, referente ao dado “*soft*”, normalmente, percepção, concordância etc.; e objetiva, cujo dado é chamado de “*hard*”, como taxas de rotatividade e de produtividade.

Com relação ao nível em que a medida foi aplicada, identificou-se que os estudos realizaram a avaliação de efetividade de GD utilizando, como apontado, as medidas de nível individual por meio de consulta predominantemente realizada com múltiplas fontes, como funcionários, gestores, diretos etc. Segundo Roberts (1996), múltiplos perfis de participantes estão atrelados à tentativa de abordar tanto a perspectiva dos avaliadores quanto dos avaliados.

Entretanto, destaca-se que a percepção de satisfação do público-alvo (Impacto societal) não foi avaliada diretamente por nenhuma consulta realizada com o público-alvo (*e.g.* cidadão ou consumidores), mas, sim, a partir da percepção dos próprios autores dos estudos, como por observação ou análise documental, acerca dos dados obtidos. A predominância dos múltiplos perfis utilizados como fontes de informação demonstra que os estudos buscam abarcar diferentes perspectivas, não demonstrando haver a necessidade de consulta a um grupo em específico, a não ser quando este é o próprio objeto de estudo da pesquisa (*e.g.* gestor da unidade de GP, em Dewettinck, 2008), destacando a questão de o construto “efetividade” ser experienciado de diferentes formas a depender do ator organizacional (Colakoglu et al., 2006; Ikramullah et al., 2016).

Logo, identificou-se que os estudos realizaram a avaliação de efetividade de GD utilizando quase em sua totalidade medidas de nível individual e de natureza subjetiva (*soft*), exemplificados em itens como “em que medida os funcionários percebem que o sistema de GD é justo” (Sharma et al., 2016, 2021, utilizando Colquitt, 2001) ou “em que medida o gestor concorda que o sistema de GD aumenta a produtividade” (Longenecker et al., 1998). Portanto, como resultado, de certa forma, da quantidade mais elevada de estudos que adotam tais dimensões, a adoção de escalas ou instrumentos para avaliar a efetividade de GD fica mais evidente nas dimensões cujos fenômenos são de nível individual.

Portanto, nas subseções seguintes, serão apresentadas as dimensões para discutir seus respectivos critérios, abordando as medidas utilizadas pelos estudos revisados para avaliar a efetividade do sistema de GD.

2.3.1.1 Dimensão de Reação

A dimensão Reação foi a mais adotada para avaliação de efetividade do sistema de GD, identificada em 41 estudos, conforme descrito na Tabela 3, representando 72% da amostra revisada. Essa dimensão contempla as impressões e opiniões de atores organizacionais acerca do sistema de GD (Keeping & Levy, 2000; Levy & Williams, 2004; Schleicher et al., 2019).

Tabela 3

Estudos revisados que adotaram a dimensão Reação

Dimensão	Estudos revisados	Frequência	
		n	%
Reação	Abdullah e Abraham (2012); Awan et al. (2020); Boroujerdi et al. (2016); Cislighi e Zugno (2019); Costigan et al. (2009); Culbertson et al. (2013); Da Luz e Figueiredo Filho (2018); Dewettinck e van Dijk (2013); Dhas et al. (2015); Govender e Bussin (2020); Ikramullah et al. (2016); Iqbal et al. (2015); Iqbal et al. (2019); Ishaq et al. (2013); Kakkar et al. (2020); Keeping e Levy (2000); Kim (2016); Kivipöld et al. (2020); Lawler (2003); Levy e Williams (2004); Makhubela et al. (2016); Maley e Moeller (2014); Modipane et al. (2019); Mok e Leong (2021); Phuong (2018); Pinto e Behr (2015); Ponmuthumari (2020); Roberts (1992, 1996); Rolle e Klingner (2012); Schleicher et al. (2019); Sharma et al. (2016, 2021); Soltani et al. (2003); Tuytens e Devos (2012); Upadhyay et al. (2020); Van Waeyenberg e Decramer (2018); Wright e Cheung (2007); Yang e Hsieh (2007); Yu et al. (2018)	41	72

Essa mesma predominância em avaliar a efetividade de GD pela reação de funcionários e gestores, identificada nos estudos revisados, já havia sido relatada anteriormente por Schleicher et al. (2019) em seu estudo sobre efetividade do sistema de GD, assim como os critérios relativos à reação, os quais também aparecem como os mais adotados para avaliar a efetividade de GD. Os autores atribuem esta atual importância como resposta ao apelo feito por Murphy e Cleveland em 1995 para que os critérios de reação, considerados à época como “os critérios negligenciados” pela literatura de GD, fossem considerados para avaliação dos resultados de GD e sua efetividade.

Até então, o que se tinha era um foco predominante em critérios de efetividade do sistema de GD voltados para as propriedades da avaliação, como a psicometria e a precisão do instrumento, o que ignorava amplamente ou completamente outros critérios de interesse dos praticantes (Keeping & Levy, 2000). Essas características mantiveram, durante certo tempo, pesquisadores e praticantes distantes de critérios “*soft*”: percepções sobre o sistema no que tange à satisfação, utilidade e justiça, chamadas de reações; esses critérios dizem respeito a elementos de ordem mais prática, o que interessa mais aos praticantes (Murphy & Cleveland, 1995; Keeping; Levy, 2004). Em meio a isso, têm-se, à época, estudos apontando a insatisfação dos funcionários com relação ao sistema de GD, o que manifestou a dificuldade de aceitação do sistema, mesmo quando os padrões psicométricos se mostravam adequados (*e.g.* Evans; McShane, 1988; Longenecker et al., 1988). Logo, Keeping e Levy (2000) apontam que vários

autores passam a sugerir que as reações à avaliação desempenham um papel importante na avaliação de efetividade, pois se mostram vitais para a aceitação e o uso adequado do sistema (e.g. Cardy; Dobbins, 1994; Murphy & Cleveland, 1995), assim como um importante indicador sobre a validade da avaliação realizada (Lawler, 1967).

Em complemento, o desenvolvimento da Teoria da Troca Social (*Social Exchange Theory* – SET) é utilizada por alguns autores (e.g. Levy & Williams, 2004; Sharma et al., 2016, 2021) para justificar a importância das reações, pois, diante do contexto social em que as avaliações se inserem, os autores entendem que a forma com uma prática ou política organizacional é vista pode determinar outros aspectos organizacionais, como o nível de participação, engajamento e comprometimento com a ferramenta gerencial, podendo, ou não, interferir na qualidade dos resultados esperados (Pichler, 2012). Além dessa teoria, destaca-se também que os estudos utilizaram as teorias da equidade, da expectativa, do estabelecimento de metas, todas de base teórica psicológica e proposta por psicólogos (e.g. Victor Vroom, J. Stacy Adams e Edwin A. Locke), para embasar a utilização dos critérios descritos a seguir como elementos importantes das reações para avaliação de efetividade de GD. Essas teorias deixam evidente a linha predominantemente psicológica com que as reações traduzem os resultados de GD pela perspectiva dos indivíduos.

Nesse sentido, o conjunto desses apontamentos é visto por advogar em prol da importância de se considerar as reações dos funcionários no que tange ao sistema de GD (Keeping & Levy, 2000). Não à toa, nos anos seguintes, sobretudo a partir de 2000, diversos estudos passaram a considerar os critérios de reação como uma dimensão essencial para a avaliação de efetividade de GD (e.g. Levy & Williams, 2004; Keeping & Levy, 2000; Ishaq et al., 2013).

Com isso, foi possível perceber que alguns autores revisados buscaram aprofundar o conteúdo específico dessa dimensão, desenvolvendo, teoricamente e empiricamente, estudos que exploraram as reações enquanto elemento principal para avaliar a efetividade do sistema de GD (e.g. Awan, 2020; Iqbal et al., 2019; Keeping & Levy, 2000; Kim, 2016; Modipane et al., 2019; Sharma et al., 2016, 2021; Tuytens & Devos, 2012; Wright & Cheung, 2007). Alguns desses estudos utilizaram apenas a dimensão Reação para avaliar a efetividade de GD ao buscar examinar e aprofundar o estudo sobre a mensuração dessa dimensão (e.g. Keeping & Levy, 2000; Modipane et al., 2019; Wright & Cheung, 2007) ou por razões circunstâncias, que impedem a coleta de outras dimensões, tornando a dimensão de Reação a única opção para mensuração dos resultados do sistema de GD (e.g. Cislighi et al., 2019).

Diante desse resultado, percebe-se que é possível aferir a efetividade do sistema de GD utilizando somente as reações que os participantes possuem acerca do sistema. Entretanto, ressalta-se que, assim como um sistema de GD pode ser percebido como efetivo mesmo quando possui erros ou não atinge os objetivos esperados, a adoção apenas dos critérios de Reação pode não ser representativa da efetividade de GD, uma vez que sistemas percebidos como ruins ou errados ainda poderão ter notados efeitos do sistema de GD em outros níveis ou outras dimensões. Portanto, considerar apenas a dimensão Reação representa entregar uma avaliação parcial dos resultados de GD.

Identificaram-se cinco critérios da dimensão Reação que foram adotados pelos estudos revisados: Justiça, Precisão, Satisfação, Utilidade e Afetividade. Esses critérios foram identificados a partir dos estudos revisados e corroborados pela categorização oriunda de estudos correlatos, mas que não foram incluídos na revisão por não atenderem aos critérios do protocolo (*e.g.* Pichler, 2012; Murphy & Cleveland, 1995; Cawley et al., 1998). Os critérios de Reação identificados, assim como os respectivos estudos que adotaram esses critérios, estão dispostos na Tabela 4.

Tabela 4

Crítérios da dimensão Reação

Crítérios	Autores	Frequência	
		n	%
Justiça	Abdullah e Abraham (2012); Awan et al. (2020); Boroujerdi et al. (2016); Dewettinck e van Dijk (2013); Iqbal et al. (2015); Iqbal et al. (2019); Ishaq et al. (2013); Kakkar et al. (2020); Keeping e Levy (2000); Kim (2016); Kivipöld et al. (2020); Levy e Williams (2004); Makhubela et al. (2016); Modipane et al. (2019); Mok e Leong (2021); Phuong (2018); Roberts (1992); Roberts (1996); Schleicher et al. (2019); Sharma et al. (2016); Sharma et al. (2021); Tuytens e Devos (2012); Van Waeyenberg e Decramer (2018)	24	42
Precisão	Awan et al. (2020); Costigan et al. (2009); Da Luz e Figueiredo Filho (2018); Ikramullah et al. (2016); Iqbal et al. (2015); Kakkar et al. (2020); Keeping e Levy (2000); Lawler (2003); Levy e Williams (2004); Longenecker et al. (1988); Modipane et al. (2019); Pinto e Behr (2015); Roberts (1996); Rolle e Klingner (2012); Schleicher et al. (2019); Sharma et	20	35

	al. (2016); Sharma et al. (2021); Van Waeyenberg e Decramer (2018); Wright e Cheung (2007); Yang e Hsieh (2007)		
Satisfação	Abdullah e Abraham (2012); Culbertson et al. (2013); Dhas et al. (2015); Govender e Bussin (2020); Iqbal et al. (2015); Iqbal et al. (2019); Ishaq et al. (2013); Keeping e Levy (2000); Levy e Williams (2004); Mok e Leong (2021); Pinto e Behr (2015); Ponmuthumari (2020); Schleicher et al. (2019); Upadhyay et al. (2020); Van Waeyenberg e Decramer (2018)	14	25
Utilidade	Abdullah e Abraham (2012); Boroujerdi et al. (2016); Cislighi e Zugno (2019); Dhas et al. (2015); Keeping e Levy (2000); Lawler (2003); Levy e Williams (2004); Maley e Moeller (2014); Pinto e Behr (2015); Schleicher et al. (2019); Soltani et al. (2003); Tuytens e Devos (2012); Yu et al. (2018);	13	23
Afetividade	Biulchi e Pauli (2012); Costigan et al. (2009); Da Luz e Figueiredo Filho (2018); Keeping e Levy (2000); Mok e Leong (2021); Phuong (2018); Pinto e Behr (2015); Schleicher et al. (2019)	8	14

O critério mais adotado da dimensão Reação foi Justiça, o qual consiste na extensão com que os funcionários percebem o sistema de GD como justo e igualitário (Sharma et al., 2016). Esse critério foi adotado em 24 estudos revisados, o que equivale a 42% do total dos estudos revisados.

Os autores que utilizam esse critério para avaliar a efetividade de GD se baseiam nas teorias do construto “justiça organizacional”, discorrendo sobre a importância de os funcionários perceberem processos organizacionais como justos, conforme apontado em estudos seminais de autores como Jerald Greenberg e Jason A. Colquitt, referenciados pelos estudos revisados (*e.g.* Abdullah & Abraham, 2012; Modipane et al., 2019). Nesse sentido, os autores seminais promoveram discussões sobre como o conceito de justiça poderia ser mais bem conceituado e mensurado e chegaram a quatro fatores distintos, mas ainda altamente correlacionados: distributivo, procedimental, interpessoal e informacional (Sharma et al., 2016).

O primeiro fator se refere à justiça distributiva referente ao quão justo são percebidos os resultados alcançados por meio da GD, como recompensas e gratificações (Keeping & Levy, 2000; Kivipöld et al., 2020; Iqbal et al., 2015; Modipane et al., 2019; Sharma et al., 2016, 2020; Phuong, 2018; Tuytens & Devos, 2012). A justiça distributiva é conceituada com base na Teoria da Equidade (*equity theory*), de Adams (1963), ao considerar o nível com que os funcionários

percebem o quão justo são resultados e as recompensas e gratificações distribuídas (Kivipõld et al., 2020; Modipane et al., 2019; Sharma et al., 2016).

A justiça processual envolve a percepção de participação e da possibilidade de opinar no processo e nos procedimentos decisórios envolvidos no sistema de GD (Abdullah; Abrahm, 2012; Kivipõld et al., 2020; Modipane et al., 2019; Sharma et al., 2016). No estudo de Modipane et al. (2019), os autores identificaram estudos que destacam a capacidade de a justiça procedimental influenciar a satisfação no trabalho.

A justiça interpessoal é referente ao tratamento justo nas relações interpessoais (Sharma et al., 2016) com dignidade, respeito e cortesia (Modipane et al., 2019). Esse fator é percebido conforme estudo identificado em Modipane et al. (2019) ao relatar que, quando as classificações de desempenho dos funcionários foram inferiores ao esperado, suas atitudes em relação aos resultados de GD variaram dependendo do tratamento interpessoal que receberam do avaliador (Modipane et al., 2019).

Por último, a justiça informacional, a qual se refere ao acesso a explicações e demais informações pertinentes à GD (Mahbuela et al., 2016; Modipane et al., 2019; Sharma et al., 2016). Esse fator é apontado por se relacionar com aspectos como a comunicação entre funcionários e gestores, motivação e comprometimento com o trabalho ao terem um sentimento de orgulho e excitação no que eles fazem e podem fazer (Modipane et al., 2019).

Os estudos revisados operacionalizaram o critério justiça sob esses quatro fatores (*e.g.* Abdullah; Abrahm, 2012; Keeping & Levy, 2000; Kivipõld et al., 2020; Iqbal et al., 2015; Modipane et al., 2019; Sharma et al., 2016, 2020; Phuong, 2018; Tuytens & Devos, 2012).

Alguns estudos fazem a relação da percepção de justiça com outras variáveis do sistema de GD. É o caso do estudo de Tuytens e Devos (2012) e Abdullah e Abraham (2012), os quais identificaram que a percepção de que o sistema de GD é justo auxilia na aceitação do sistema. O entendimento de que um sistema de GD justo pode resultar em uma aceitação mais elevada do sistema por parte dos funcionários é apontado por resultar em maior comprometimento organizacional e satisfação com o trabalho (Abdullah & Abraham, 2012).

Selden e Sowa (2011), em seu estudo sobre organizações do terceiro setor, identificaram que a percepção de um sistema de GD justo está atrelado ao fornecimento de feedback úteis. O fornecimento de feedback útil está relacionado com um sistema de GD justo, acompanhado de treinamento efetivo dos funcionários e envolvimento dos superiores, o que, segundo descrito por Abdullah e Abraham (2012), garante satisfação com o trabalho.

No estudo de Dewettinck e van Dijk (2013) sobre organização do setor público da Holanda, os autores apontam o critério justiça como essencial para a efetividade do sistema de

GD. Dewettinck e van Dijk (2013) e Schleicher et al. (2019) complementam os achados ao afirmar que a percepção de justiça enquanto determinante da efetividade de GD é mais influenciada por aspectos informais do sistema (i.e. momentos informais de feedback) do que por elementos previstos formalmente ou de forma estruturada no sistema de GD.

Nos últimos anos, o critério justiça tem assumido um papel relevante na avaliação de efetividade do sistema de GD ao ser adotado em conjunto com o critério precisão, descrito a seguir. Essa forma de avaliação foi cunhada por Sharma et al. (2016), que consideram a percepção de efetividade de GD – ou chamada de “*perceived performance management system effectiveness*” (PMSE) – e indicam que a percepção exclusiva sobre precisão e justiça seria suficiente para avaliar a efetividade de GD. Desde então, percebe-se uma tendência por parte dos estudos em considerar somente esses dois critérios para avaliação de efetividade de GD, como Awan et al. (2020), Kakkar et al. (2020), Modipane et al. (2019) e Sharma et al. (2020).

Nesse sentido, o segundo critério mais adotado da dimensão de Reação foi o de Precisão, o qual refere-se à percepção do sistema de GD acerca da capacidade de refletir corretamente aquilo a que ele se propõe a fazer (Sharma et al., 2016). Esse critério foi utilizado por 20 artigos (35%).

Para aferir essa precisão, os estudos adotam elementos distintos. Um desses elementos refere-se à avaliação de erros e vieses que acometem o sistema de GD, adotada em Iqbal et al. (2015), os quais fazem a diferenciação de erros em quatro grupos: i) erros centrados no avaliador, relacionados a aspectos demográficos (idade, gênero e nível educacional) e variáveis psicológicas (autoconfiança, autoestima, personalidade); ii) no avaliado, a partir de seus comportamentos, exemplificado pelo autor como impressionar o avaliador, insinuações ou buscar melhorar sua reputação; iii) no próprio sistema, como políticas, procedimentos e suporte, incluindo instrumento de avaliação ou a funcionalidade do sistema informatizado de GD; ou iv) nas relações avaliador-avaliado, destacada pelos autores como aspectos éticos, relacionados a questão racial ou política.

Outra classificação utilizada é a dos autores Sharma et al. (2016), os quais desenvolveram uma escala para mensurar a percepção de precisão que está ancorada em quatro fatores: i) o planejamento do desempenho, avaliando o quanto os funcionários percebem que os desempenhos individuais planejados estão baseados nos objetivos organizacionais (precisão do planejamento do desempenho); ii) o feedback e o aconselhamento, relacionado ao quanto os funcionários percebem que os desempenhos individuais apresentados estão baseados naquilo que foi planejado, indicando por meio de feedbacks e aconselhamentos regulares; iii) a revisão do desempenho, avaliando o quanto os desempenhos individuais anuais apresentados foram

revisados para se manter alinhado com o que foi planejado; iv) os resultados, os quais se referem ao quanto os resultados de GD, no que tange à classificação, remuneração, recompensa e/ou reconhecimento, estão baseados no desempenho individual apresentado.

Em outras palavras, Awan et al. (2020) afirmam que esses fatores de precisão do sistema de GD refletem de forma combinada as teorias apontadas por DeNisi e Pritchard (2006) e Dewettinck e Van Dijk (2013) como justificativas sobre como o sistema de GD funciona como uma ferramenta gerencial útil e promissora para melhoria do desempenho, como a da expectativa (*expectancy theory*, Vroom, 1964), a qual se baseia na suposição de que os indivíduos tendem a alocar quantidade limitada de tempo e energia para ações das quais eles esperam que as consequências maximizem sua satisfação (DeNisi & Pritchard 2006); a do desenvolvimento de metas (*goal-setting theory* – GST; Locke et al., 1981), a qual foi desenvolvida para explicar como as metas de desempenho afetam o desempenho de tarefa e considera o conteúdo das metas de desempenho como um fator determinante em quão bem a tarefa é executada, em que os autores Locke e Latham (2002) identificaram a mediação de feedback regular entre os resultados atuais e os resultados esperados; e de Controle (*control theory*; Klein, 1989), que diz respeito ao processo que reduz a diferença entre as consequências das ações atuais e dos resultados esperados (Donovan, 2001; Buchner, 2007).

A escala de precisão do sistema de GD desenvolvida por Sharma et al. (2016) foi utilizada por outros estudos revisados, como os dos autores Awan et al. (2020), Kakkar et al. (2020) e Modipane et al. (2019), para avaliar a efetividade de GD. Os próprios autores que desenvolveram a escala a utilizaram recentemente em outro estudo (cf. Sharma et al., 2021) em resposta ao estudo desenvolvido pelos autores Kakkar et al. (2020) para avaliar a mesma relação estabelecida em um modelo de avaliação pelos autores. Porém, nesse caso, Sharma et al. (2021) utilizaram a combinação de justiça e precisão para avaliar a efetividade de GD e não somente a precisão, como feito pelos autores Kakkar et al. (2020).

Logo, percebe-se certa convergência entre os estudos em adotar a escala de Sharma et al. (2016) para mensurar a percepção de precisão enquanto critério da avaliação de efetividade de GD, uma vez que nenhuma outra escala de outro estudo foi adotada de forma recorrente entre os estudos revisados como acontece com a de Sharma et al. (2016).

Em consonância com os fatores descritos por Sharma et al. (2016) e Iqbal et al. (2013), destaca-se que outros autores também adotaram elementos relacionados para mensurar o critério precisão como: facilidade e clareza em compreender o que está sendo avaliado, adotado pelo estudo de Van Waeyenberg e Decramer (2018); a capacidade de avaliar corretamente o desempenho, adotada pelo estudo de Rolle e Klingner (2012) e Wright e Cheung (2007); no

estudo de Ikramullah et al. (2016) e de Yang e Hsieh (2007), avaliou-se a efetividade de GD a partir da identificação de erros ou vieses dos avaliadores durante o processo de avaliação, considerando aspectos como a validade e confiabilidade da avaliação no sistema de GD como medidas de precisão; e utilizaram os elementos de complexidade e subjetividade relacionados à avaliação de desempenho do sistema de GD, como no caso do item no questionário de Longenecker et al. (1998) em que se indaga ao funcionário e ao gestor se a “AD é muito subjetiva”.

O terceiro critério mais adotado foi o de Satisfação, aplicado em 14 artigos revisados (25%), definido como o nível de contentamento geral com o sistema de GD (Culbertson et al., 2013; Schleicher et al., 2019). Segundo Keeping e Levy (2000), o critério Satisfação possui uma natureza tipicamente genérica, o que resulta em operacionalizações inconsistentes e contaminadas por outras variáveis. Por essa razão, questionou-se, inicialmente, a pertinência dessa categoria enquanto critério de avaliação de efetividade de GD neste trabalho com receio de uma eventual sobreposição com os demais critérios, o que ocasionaria na violação de parte das regras de análise de conteúdo.

Entretanto, este estudo considerou relevante manter o critério Satisfação como uma das categorias por ter sido assim identificado nos estudos revistados, sobretudo como satisfação geral com o sistema de GD. Além disso, Keeping e Levy (2000) mantiveram esse critério separado por afirmarem que os critérios de reação de GD são mais bem trabalhados quando figurados de forma separada; de forma semelhante, Schleicher et al. (2019) também optaram por manter esse critério em separado como parte do modelo de avaliação de efetividade de GD. Além disso, segundo Culbertson et al. (2013), a satisfação com GD é considerada uma das categorias de reação que mais indica as consequências ao feedback referente a um processo de avaliação, tendo em vista a relação estabelecida para aumentar motivação, comprometimento e desempenho.

Nesse critério foram identificados os estudos que avaliaram a efetividade do sistema de GD a partir da satisfação de forma genérica ou sem pormenorizar, restringindo-se a “satisfação com o sistema de GD” com itens referentes a “estou satisfeito com a forma que desempenho é avaliado” (*e.g.* Levy & Williams, 2004; Pinto & Behr, 2015; Upadhyay et al., 2020; Van Waeyenberg & Decramer, 2018). Nessa abordagem é possível perceber que a operacionalização desse critério é, tal como colocado por Keeping e Levy (2004), inconsistente e potencialmente confusa, uma vez que deixa em aberto aspectos que deveriam ser considerados, fazendo com que o critério possa se sobrepor a outros.

Por outro lado, alguns estudos fizeram a identificação do que deveria ser considerado com relação à satisfação do sistema de GD, especificando a satisfação com: o instrumento de avaliação, as recompensas, os resultados ou a sessão de feedback etc. (*e.g.* Abdullah & Abraham, 2012; Dhas et al., 2015; Culbertson et al., 2013; Iqbal et al., 2015; Ponmuthumari, 2020).

Não foi identificada uma predominância por medidas específicas no que tange à avaliação da Satisfação em relação ao sistema de GD. Pontmuthumari (2020) e Upadhyay et al. (2020) relataram que os funcionários se sentem satisfeitos em relação a como a avaliação de desempenho é realizada. Culbertson et al. (2013) e Pontmuthumari (2020) relataram que os funcionários se sentem satisfeitos em relação ao feedback fornecido com qualidade. Abdullah e Abraham (2012) também identificaram que os funcionários estão satisfeitos com o sistema de GD em relação à clareza e aos objetivos da avaliação de desempenho, expressos de forma clara, assim como a forma como os resultados de GD são transmitidos e utilizados pela organização. Por outro lado, nos achados de Govender e Bussin (2020) e Pinto e Behr (2015), foi apontada a insatisfação dos funcionários com relação à forma como o sistema de GD é conduzido.

O penúltimo critério mais adotado foi o de Utilidade, aplicado em 13 artigos revisados (23%), com intuito de avaliar a capacidade do sistema de GD em fornecer insumos para outras políticas e práticas da organização (Ikramullah et al., 2016). Essa categoria, comparada às demais de reação, é considerada mais consistente e menos confusa (Keeping & Levy, 2000).

Dessa forma, os artigos avaliaram a efetividade do sistema de GD considerando a percepção de utilidade baseada no quanto a ferramenta atende seu propósito (*e.g.* Abdullah & Abraham, 2012) e a sua importância para o alcance de resultados (*e.g.* Cislighi & Zugno, 2019; Keeping & Levy, 2000). Outro ponto explorado pelos estudos consiste na utilidade do feedback recebido (*e.g.* Boroujerdi et al., 2016; Lawler, 2003; Soltani et al., 2003; Tuytens & Devos, 2012; Yu et al., 2018).

Os estudos identificaram a percepção de que a GD é útil como ferramenta para progressão individual na organização (*e.g.* Abdullah & Abraham, 2012; Dhas et al., 2015; Pinto; Behr, 2015). Outros artigos relatam que os funcionários reconhecem o feedback fornecido pelo sistema de GD como útil para prever e auxiliar no desempenho futuro (*e.g.* Boroujerdi et al., 2016; Lawler, 2003; Soltani et al., 2003; Tuytens & Devos, 2012; Yu et al., 2018).

Por último, identificou-se o critério Afetividade, adotado em oito artigos revisados (14%), referente aos sentimentos experienciados pelos usuários em relação ao sistema de GD (Schleicher et al., 2019). Keeping e Levy (2000) relatam a presença de efeitos positivos ao funcionário em relação ao sistema de GD, como felicidade, otimismo, empatia, orgulho,

confiança e prazer, e de efeitos negativos, como raiva, ansiedade, irritação e agitação. Também foi identificado o sentimento de confiança, tanto no processo do sistema de GD (*e.g.* Mok & Leong, 2021), como na capacidade de avaliação do avaliador (*e.g.* Phuong, 2018); felicidade, no estudo de Abu-Shanab et al. (2021); e o sentimento de se sentir reconhecido (*e.g.* Costigan et al., 2009; Luz & Figueiredo, 2018).

Já com relação aos efeitos negativos, o sistema de GD provocou, no estudo de Pinto e Behr (2015), o sentimento de receio em avaliar a chefia imediata, como também foi apontado por Ishaq et al. (2009) como um catalisador de angústia emocional para os funcionários. De forma semelhante, Biulchi e Pauli (2012) afirmam que a relação do sistema de GD como fator variável na remuneração recebida é percebida pelos funcionários como um fator de insegurança financeira.

Em resumo, identificou-se que a dimensão de Reação é mensurada a partir da adoção de cinco critérios distintos, sendo que Justiça foi o mais identificado por ter sido adotado entre os estudos e Afetividade com a menor recorrência. Assim, a efetividade do sistema de GD pode ser aferida por critérios de reação que são adotados tanto de forma isolada (*e.g.* satisfação ou utilidade) como de forma combinada entre si, isto é, a mensuração emprega mais de um, destacando, principalmente, a combinação entre os critérios Justiça e Precisão, tendo em vista a sua preferência conferida por certos estudos recentes.

Após a identificação dos critérios, considerou-se válido identificar os padrões metodológicos adotados. Nesse sentido, com relação a como a coleta foi realizada pelos estudos que adotaram os critérios da dimensão de Reação, identificou-se uma predominância pela abordagem quantitativa (63%) com dados de origem primária (80%) obtidos por meio de questionários (61%) (*e.g.* Awan et al., 2020; Costigan et al., 2009; Culberston et al., 2013; Dewettinck & van Dijk, 2013; Dhas et al., 2015; Iqbal et al., 2019; Ishaq et al., 2013; Kakkar et al., 2020; Keeping & Levy, 2000; Kim, 2016; Lawler, 2003; Makhubela et al., 2016; Modipane et al., 2019; Mok & Leong, 2021; Phuong, 2018; Pomuthumari, 2020; Roberts, 1992; Roberts, 1996; Sharma et al., 2021; Soltani et al., 2003; Tuytens & Devos, 2012; Upadhyay et al., 2020; Van Waeyenberg & Decramer, 2018; Yang & Hsieh, 2007; Yu et al., 2018). Esses questionários foram aplicados com medidas de nível individual, utilizando somente a percepção, principalmente, de múltiplos perfis das organizações (49%), abrangendo diferentes níveis organizacionais (*e.g.* funcionários, gestores de linha e diretores). Os múltiplos perfis de participantes dos estudos estão atrelados à tentativa de abordar tanto a perspectiva dos avaliadores quanto dos avaliados (*e.g.* Roberts, 1996).

Tendo em vista a ausência de informações consolidadas sobre a mensuração da efetividade de GD demonstrada até aqui, considerou-se relevante identificar se havia um padrão nos instrumentos de pesquisa adotados pelos estudos revisados, principalmente na dimensão de Reação, que, por haver uma maior quantidade de estudos, permitia identificar a existência de homogeneidade nos instrumentos de coleta adotados.

Nesse sentido, identificou-se que os instrumentos de pesquisas variam a depender do critério utilizado pelo estudo, mas destaca-se o instrumento de Colquitt (2001), para mensuração do critério de Justiça, e o de Sharma et al. (2016), para avaliar a percepção da precisão do sistema de GD, uma vez que foram os instrumentos empregados de forma mais recorrente que os demais instrumentos utilizados. É possível perceber uma certa convergência na mensuração dos critérios de justiça e precisão em conjunto para aferir a efetividade de GD, principalmente em estudos mais recentes (*e.g.* Awan et al., 2020; Modipane et al., 2019; Sharma et al., 2016, 2021).

Os critérios de Satisfação e Utilidade também apresentam escalas passíveis de aplicação e adaptação em outros contextos a partir da Escala de Satisfação com o Sistema de GD de Decramer et al. (2012) (Van Wayenberg & Decramer, 2018). Já para o critério de Utilidade, destaca-se a medida de percepção da utilidade do feedback da escala de Heneman e Milanowski (2003), adotada por Tuytens e Devos (2012).

Assim, conclui-se que essa dimensão é predominantemente pautada na percepção dos envolvidos acerca do sistema de GD. Predominância esta que se assenta tanto na natureza do fenômeno abarcado por essa dimensão, como na conveniência de acesso a esse tipo de informação por parte das unidades de GP.

Considerando que esses dois fatores também justificam a adoção elevada dessa dimensão por parte dos estudos revisados e que isso pode indicar uma aparente relevância dos critérios e da dimensão de Reação para avaliar a efetividade de GD, a revisão também demonstrou ser ainda latente a necessidade de maior clareza na definição constitutiva de alguns critérios a fim de sanar confusões conceituais, como apontaram Murphy e Cleveland ainda em 1995. Logo, torna-se essencial que os estudos busquem corrigir tais problemas conceituais em razão da predominância de estudos que adotam essa dimensão, principalmente para estudos que adotam exclusivamente ou que consideram essa dimensão como essencial para avaliar a efetividade de GD (*e.g.* Cislighi & Zugno, 2019; Costigan et al., 2009; Dhas et al., 2015; Govender & Bussin, 2020; Iqbal et al., 2015; Keeping & Levy, 2000; Kim, 2016; Makhubela et al., 2016; Modipane et al., 2019; Sharma et al., 2016, 2021; Tuytens & Devos, 2012; Wright & Cheung, 2007), evitando problemas de confiabilidade e validade metodológica.

Apesar dos aspectos que advogam a favor dessa dimensão, Schleicher et al. (2019) apontam que a sua mensuração não é tida como suficientemente capaz de determinar que um sistema de GD é ou não fracassado, ainda que ele seja percebido como injusto, impreciso, insatisfatório, inútil e traga sentimentos e efeitos ruins. Em geral, Murphy (2020) aponta que entre as organizações, apesar da maioria de seus funcionários não aprovar o sistema de GD, isso não é fator suficiente para abandonar a GD. O autor ainda complementa taxativamente que, mesmo com a preocupação excessiva em evitar erros e tornar o sistema justo e útil, não é possível garantir os resultados desejados, pelo contrário: ele afirma que os sistemas de GD sempre serão cheios de erros, inúteis em certa medida e injustos.

2.3.1.2 Dimensão de Aprendizagem

A dimensão Aprendizagem foi a terceira mais adotada para avaliação de efetividade do sistema de GD, figurando em 31 artigos, conforme descrito na Tabela 5, representando 54% da amostra revisada. Essa dimensão reflete os conhecimentos, habilidades e atitudes adquiridos pelo indivíduo a partir da participação no sistema de GD (Schleicher et al., 2019), seja ao final de um ciclo ou no decorrer do processo.

Tabela 5

Estudos revisados que adotaram a dimensão Aprendizagem

Dimensão	Estudo revisado	Frequência	
		n	%
Aprendizagem	Abdullah e Abraham (2012); Ahmed (1999); Culbertson et al. (2013); Dewettinck (2008); Dewettinck e van Dijk (2013); Haines e St-Onge (2012); Ikramullah et al. (2016); Iqbal et al. (2019); Ishaq et al. (2009); Ishaq et al. (2013); Kakkar et al. (2020); Kang e Choi (2019); Lawler (2003); Levy e Williams (2004); Longenecker e Nykodym (1996); Longenecker et al. (1988); Lopes et al. (2020); Mok e Leong (2021); Phoung (2018); Ponmuthumari (2020); Roberts (1992, 1996); Rolle e Klingner (2012); Schleicher et al. (2019); Selden e Sowa (2011); Sharma et al. (2021); Teeroovengadum et al. (2019); Upadhyay et al. (2020); Van Waeyenberg e Decramer (2018); Yang e Hsieh (2007); Yu et al. (2018)	31	54

É importante destacar que os critérios identificados aqui como pertencentes à Aprendizagem figuravam, até então, de maneira isolada ou pontual, sem menção ou relação a algum grupo específico, como uma categoria ou dimensão. Mas, em 2019, Schleicher et al.

foram os primeiros – e até aqui os únicos – a condensarem esses critérios na dimensão Aprendizagem, momento em que passa a figurar dessa forma em um modelo de avaliação de efetividade de GD. Portanto, a dimensão Aprendizagem somente aparece assim recentemente, diferenciando-se, principalmente, da dimensão Reação, cuja incorporação junto à discussão da avaliação de efetividade de GD tem acontecido desde a década de 1990 (*e.g.* DeNisi & Peters, 1996; Murphy & Cleveland, 1995).

Tendo em vista a ausência da dimensão Aprendizagem em modelos avaliativos anteriores, os autores Schleicher et al. (2019) expressam certa preocupação pela aprendizagem até então não ter aparecido dessa forma em outros estudos sobre o tema da efetividade de GD e afirmam que sem ela pode não ficar claro como a GD alcança seus objetivos. Isso porque, segundo os autores, é importante diferenciar a dimensão Reação da Aprendizagem para compreender a efetividade de GD, pois, referenciando Kraiger et al. (1993), “as reações capturam como o sistema de GD é experienciado pelos funcionários, mas não são medidas diretas do que se pôde ter aprendido enquanto resultado desta experiência” (Schleicher et al., 2019, p. 857).

Em geral, esta revisão identificou que os critérios pertencentes a essa dimensão tratam dos resultados decorrentes da participação no sistema de GD: obtenção de clareza com relação às tarefas e funções desempenhadas, compreensão dos benefícios do sistema de GD, desenvolvimento de técnicas de fornecimento e recebimento de feedback mais efetivos, mudanças na intenção ou propensão de se comportar no trabalho e outros aspectos que são tidos como resultados individuais que se relacionam à GD. Logo, percebe-se que os autores buscaram explicar o mecanismo pelo qual aquelas reações se transformam em outros resultados, uma vez que somente elas não mantêm relação causal suficiente para precisar o que se pode ter adquirido como resultado da participação no sistema de GD. Portanto, mesmo que a efetividade de GD possa ser compreendida pelas reações, como há uma grande distância das reações para os demais resultados de GD, o caminho exige que seja tratada também a questão da aprendizagem.

Portanto, o reconhecimento desses critérios para avaliação de efetividade de GD nos estudos revisados se deu mais pela sua identificação como itens de questionários utilizados para avaliação de efetividade de GD enquanto medida, descritas como a capacidade do sistema de GD em “contribuir para o desenvolvimento profissional” (Dewettinck & van Dijk, 2013), desenvolver habilidades (Yu et al., 2018), “auxiliar a clarear os objetivos e metas de desempenho” (Longenecker et al., 1988), “entender e aceitar o processo de GD” (Rolle & Klingner, 2012) do que variáveis ou medidas separadas que compuseram modelos teóricos. Por

essa razão, a identificação de teorizações ou de explicações aprofundadas do porquê da adoção desses critérios por parte de alguns autores de alguns estudos ficou defasada.

Apesar disso, os critérios identificados puderam ser sistematizados nas categorias: Conhecimento, Habilidade e Atitude, as quais serão definidas e explicadas adiante, e que foram definidas de acordo com a classificação convencionada pela literatura de aprendizagem (*e.g.* Abbad et al., 2012; Kraiger & Ford, 2021), a mesma utilizada pelos autores Schleicher et al. (2019) ao sistematizarem a dimensão Aprendizagem. Portanto, os critérios adotados por esta revisão se assemelham aos criados pelos autores Schleicher et al. (2019) que, ao reconhecerem a importância dessa dimensão para o modelo de avaliação de efetividade de GD e diante da ausência de um arcabouço teórico próprio que fundamentasse a criação de suas categorias, apropriaram-se da literatura consagrada de avaliação de treinamento para propor a dimensão de Aprendizagem.

Ressalta-se que, apesar da ordem lógica normalmente estabelecida e assim apresentada pela literatura de treinamento quando se fala de Conhecimentos, Habilidades e Atitudes, o texto a seguir foi apresentado seguindo a ordem de maior recorrência entre os critérios. Assim, a Tabela 6 reúne esses critérios da dimensão Aprendizagem, assim como os estudos revisados que adotaram essa dimensão.

Tabela 6

Critérios da dimensão Aprendizagem

Critérios	Autores	Frequência	
		n	%
Atitude	Abdullah e Abraham (2012); Dewettinck (2008); Dewettinck e van Dijk (2013); Haines e St-Onge (2012); Ikramullah et al. (2016); Ishaq et al. (2013); Kakkar et al. (2020); Kang e Choi (2019); Levy e Williams (2004); Longenecker e Nykodym (1996); Longenecker et al. (1988); Phuong (2018); Roberts (1992); Roberts (1996); Rolle e Klingner (2012); Schleicher et al. (2019); Sharma et al. (2021); Upadhyay et al. (2020)	18	32
Habilidade	Ahmed (1999); Culbertson et al. (2013); Dewettinck (2008); Dewettinck e van Dijk (2013); Haines e St-Onge (2012); Ikramullah et al. (2016); Iqbal et al. (2019); Lawler (2003); Lopes et al. (2020); Mok e Leong (2021); Rolle e Klingner (2012); Schleicher et al. (2019); Selden e Sowa (2011); Teeroovengadum et al. (2019); Van Waeyenberg e Decramer (2018); Yang e Hsieh (2007); Yu et al. (2018)	17	28

Conhecimento	Abdullah e Abraham (2012); Ikramullah et al. (2016); Iqbal et al. (2019); Ishaq et al. (2009); Kang e Choi (2019); Lawler (2003); Longenecker e Nykodym (1996); Longenecker et al. (1988); Ponmuthumari (2020); Rolle e Klingner (2012); Schleicher et al. (2019); Teeroovengadum et al. (2019); Yu et al. (2018)	13 23
---------------------	---	-------

O critério mais adotado da dimensão de Aprendizagem foi a Atitude, referente à intenção de se comportar ou agir em decorrência da interação com o sistema de GD (Schleicher et al., 2019). Esse critério foi adotado por 18 estudos, correspondendo a 32% da amostra. Para Ikramullah et al. (2016), a efetividade de um sistema de GD pode ser avaliada a partir de mudanças nas ações e atitudes dos funcionários, que podem levar a melhores resultados organizacionais.

Nesse sentido, alguns estudos (*e.g.* Abdullah & Abraham, 2012; Ikramullah et al., 2016; Kang & Choi, 2019; Longenecker & Nykodym, 1996; Longenecker et al., 1988; Roberts, 1992; 1996) apontaram as atitudes participativa e colaborativa no sistema de GD como resultado a ser avaliado. Também foi avaliada a efetividade do sistema de GD em estimular a colaboração no trabalho (*e.g.* Dewettinck, 2008; Dewettinck & van Dijk, 2015). Segundo Haines e St-Onge (2012), a atitude de mencionar publicamente um trabalho realizado em meio a uma reunião ou um grupo consiste em uma ação de reconhecimento que pode indicar a efetividade do sistema de GD. Segundo Ikramullah et al. (2016), os avaliadores são incentivados a buscar a participação dos avaliados no processo de GD para que, assim, as decisões sejam tomadas da melhor forma. Para os autores, isso não apenas cria consenso, mas também diminui as chances de conflitos e, conseqüentemente, os avaliados ficam mais satisfeitos e comprometidos com sua organização.

Nos estudos de Longenecker & Nykodym (1996) e Longenecker et al. (1988) a intenção de contribuir, participar ou continuar participando da GD também é apontada como resultado da efetividade do sistema de GD, tendo sido aferida pela indicação dos avaliados que participariam da GD mesmo que ela não fosse um processo mandatório. Nesse sentido, Kakkar et al. (2020), Phuong (2018) e Sharma et al. (2021) também avaliaram a intenção do funcionário de deixar o trabalho após o sistema de GD. A intenção de continuar na organização também foi avaliada por Upadhyay et al. (2020) pela medida “retenção do funcionário”, em que se identificou, a partir da aplicação do instrumento já validado de Kyndt et al. (2009), que uma política de benefícios e de supervisão baseada no sistema de GD pode reduzir a intenção de deixar a organização.

Outro aspecto é apontado por Abdullah e Abraham (2012), Levy e Williams (2004), Roberts (1992, 1996) e Rolle e Klinger (2012) ao afirmarem que aceitar o sistema de GD é um resultado importante a ser obtido e que indicaria a efetividade do sistema. Abdullah e Abraham (2012) identificaram que a aceitação do sistema de GD é influenciada pelo quanto os envolvidos percebem o sistema como justo, o que ocasiona maior compromisso organizacional, satisfação no trabalho e melhorias na confiança no trabalho entre o superior e supervisionado, critérios considerados na dimensão seguinte de Transferência. Esse aspecto também é apontado como um elemento central na garantia de outros resultados de GD relacionados ao trabalho, como motivação, critério também descrito na dimensão seguinte de Transferência, produtividade e redução do absenteísmo, relacionados a resultados operacionais macro-organizacionais (Roberts, 1992). Logo, aceitar o sistema de GD foi relacionado pelos autores por influenciar em resultados percebidos no desempenho do seu trabalho e que podem refletir nos resultados operacionais da organização.

O segundo critério consiste na Habilidade, referente à capacidade de aplicar os conhecimentos adquiridos com o sistema de GD (Ikramullah et al., 2016; Lawler, 2003; Schleicher et al., 2019), e foi o mais adotado dessa dimensão. Esse critério foi adotado em 17 estudos revisados, o que equivale a 30% do total dos estudos revisados.

Alguns estudos especificaram as habilidades desenvolvidas a partir do sistema de GD. Nesse sentido, a habilidade mais citada refere-se ao aprimoramento em fornecer e receber feedback adequado (*e.g.* Ahmed, 1999; Culbertson et al., 2013; Ikramullah et al., 2016; Iqbal et al., 2019; Lawler, 2003; Mok & Leong, 2021; Selden & Sowa, 2011; Teerovengadum et al., 2019). Essa habilidade aparece como um dos principais resultados obtidos do sistema de GD no que tange à aprendizagem. Segundo Culbertson et al. (2013), o fornecimento de feedbacks efetivos, sobretudo quando são positivos, relacionam-se à ativação no funcionário para um comportamento mais voltado para metas e objetivos, o que pode promover uma satisfação em relação ao sistema de GD.

Também foi apontado como sendo uma das habilidades avaliadas o desenvolvimento da capacidade de autogerenciamento do funcionário de forma que ele consiga estabelecer seus próprios objetivos e metas e outros encaminhamentos à luz das necessidades da organização e em uma perspectiva proativa, fazendo uso adequado dos feedbacks fornecidos e recebidos a partir do sistema de GD (Lopes et al., 2020).

As habilidades desenvolvidas pelos gestores enquanto avaliadores também foi identificada como resultado de GD (*e.g.* Haines & St-Onge, 2012; Rolle & Klinger, 2012; Teeroobengadum et al., 2019; Van Waeyenberg & Decramer, 2018; Yang & Hsieh, 2007).

Haines e St-Onge (2012) identificam que o desenvolvimento de habilidades pelos gestores relativas ao papel desempenhado por eles no processo de GD pode garantir que os resultados do sistema de GD sejam mais facilmente alcançados, apontando que um sistema de GD é percebido com maior efetividade quando há uma formação mais extensa e aprofundada dos avaliadores. O questionário mensurou em que medida os gestores recebem treinamento sobre conteúdos relacionados ao planejamento de desempenho, entrega de feedback e outros tópicos semelhantes em que elevados índices de resposta refletem um treinamento mais extensivo sobre diferentes conteúdos do sistema de GD.

Por outro lado, alguns autores dos estudos que utilizaram esse critério avaliaram se o sistema de GD era capaz de produzir resultados relacionados ao desenvolvimento de habilidades de forma geral, sem especificar qual habilidade fora obtida. Isso porque, conforme apontado por Lawler (2003), a partir do sistema de GD seria possível desenvolver habilidades, as quais levariam ao alcance de um desempenho superior. Entretanto, nesse intento, estudos como de Lawler (2003) e Yu, Baird e Tung (2018) avaliaram a capacidade do sistema de GD em promover habilidades individuais, mas não especificam a quais habilidades desenvolvidas se referiam. Também sem especificar tais habilidades, Dewettinck (2008) avalia a efetividade do sistema de GD focado no desenvolvimento. Isso inclui estimular o desenvolvimento de habilidades e o fortalecimento de competências individuais. Esses dois mesmos critérios foram utilizados mais tarde no estudo de Dewettinck e Van Djik (2015). Portanto, percebe-se uma ausência na especificação de quais habilidades estavam sendo avaliadas, decorrente, em partes, da forma com que esse critério foi adotado pelos estudos: enquanto item no questionário para operacionalização da medida de efetividade de GD, levantando-se com os respondentes, por exemplo, “em que medida a GD tem desenvolvido habilidades” (*e.g.* Lawler, 2003).

Por último, o critério menos identificado na dimensão Aprendizagem foi o de Conhecimento, identificado em 13 estudos, correspondendo a 23% da amostra de estudos revisados, e que se refere às informações declarativas, procedimentais ou tácitas adquiridas em decorrência da interação com o sistema de GD (Schleicher et al., 2019).

Um dos principais aspectos identificados para avaliar a efetividade de GD foi fornecer clareza às informações organizacionais, isto é, às tarefas desempenhas e ao desempenho esperado pela organização (*e.g.* Abdullah & Abraham, 2012; Ikramullah et al., 2016; Iqbal et al., 2019; Ishaq et al., 2019; Longenecker & Nykodym, 1996; Longenecker et al., 1988; Ponmuthumari, 2020; Rolle; Kligner, 2012; Teeroovengadum et al., 2019).

O conhecimento e o entendimento sobre as tarefas foram fortemente impactados pelo sistema de GD. Os estudos retratam que saber o que é esperado e como deve ser feito é um fator

crucial para o funcionário, pois assim há mais propensão de atingir às expectativas (*e.g.* Ikramullah et al. 2016; Kang & Choi, 2019; Teeroovengadum et al., 2019). Quanto mais clareza sobre o processo de GD, acesso às informações, o propósito da ferramenta, os padrões adotados e os resultados de desempenho, mais propenso ele estará a desenvolver uma atitude participativa em relação ao sistema de GD (Abdullah & Abraham, 2012).

O último aspecto identificado referente ao critério de Conhecimento foi apontado nos estudos de Lawler (2003) e de Yu et al. (2018) – este último utilizou o instrumento de coleta adotado em Lawler (2003) –, os quais adotaram o desenvolvimento de conhecimentos em geral, mas não especificaram quais conhecimentos foram avaliados como resultado do sistema de GD.

Com relação a como a coleta foi realizada pelos estudos que adotaram os critérios da dimensão Aprendizagem, identificou-se uma predominância pela abordagem quantitativa (77%), com dados de origem primária (84%) obtidos, principalmente, por meio de questionários (74%) (*e.g.* Abdullah e Abraham, 2012; Ahmed, 1999; Culbertson et al., 2013; Dewettinck, 2008; Dewettinck & van Dijk, 2013; Haines & St-Onge, 2012; Iqbal et al., 2019; Ishaq et al., 2009; Ishaq et al., 2013; Kakkar et al., 2020; Kang & Choi, 2019; Lawler, 2003; Longenecker & Nykodym, 1996; Longenecker et al., 1988; Lopes et al., 2020; Mok & Leong, 2021; Phuong, 2018; Ponmuthumari, 2020; Roberts, 1992; Roberts, 1996; Selden & Sowa, 2011; Sharma et al., 2021; Teeroovengadum et al., 2019; Upadhyay et al., 2020; Van Waeyenberg & Decramer, 2018; Yang & Hsieh, 2007), aplicados a nível do indivíduo, utilizando somente a percepção, principalmente, de múltiplos perfis das organizações (52%), abrangendo diferentes níveis organizacionais (*e.g.* funcionários, gestores de linha e diretores). Nesse sentido, identificou-se que as mesmas características metodológicas identificadas na dimensão de Reação foram as apresentadas na dimensão de Aprendizagem. Logo, novamente se conclui que essa dimensão é coletada majoritariamente pela percepção dos envolvidos no sistema de GD.

No caso da dimensão Aprendizagem, o critério Atitude, além de ter sido o mais adotado, destaca-se por possuir uma medida aplicada por três estudos referente à intenção de continuar ou de deixar a organização, mensurada a partir de duas escalas distintas, sendo uma delas o instrumento de *Michigan Organizational Assessment Questionnaire*, de Bowling e Hammond (2008), aplicada tanto por Kakkar et al. (2020), como por Sharma et al. (2021).

Em resumo, identificou-se que a dimensão de Aprendizagem é mensurada a partir da adoção de três critérios distintos, sendo que Atitude é o mais adotado, com a maior recorrência entre os estudos, e Conhecimento, com a menor. Conclui-se que a dimensão de Aprendizagem se mostra importante por ser a primeira a realmente identificar os resultados mais diretos e imediatos do sistema de GD, porém identificou-se ainda certa necessidade de consolidação

dessa dimensão para avaliação de efetividade do sistema de GD, assim como apontado pelos propulsores dessa dimensão na literatura (Schleicher et al., 2019).

2.3.1.3 Dimensão de Transferência

A dimensão de Transferência foi a segunda mais adotada pelos artigos revisados, figurando em 40 estudos e correspondendo a 68% da amostra revisada, conforme a Tabela 7. Identificou-se que os estudos revisados avaliam o quão capaz é o sistema de GD de transferir seus resultados para o trabalho (Schleicher et al., 2019).

Tabela 7

Estudos revisados que adotaram a dimensão Transferência

Dimensão	Estudos revisados	Frequência	
		n	%
Transferência	Abdullah e Abraham (2012); Ahmed (1999); Awan et al. (2020); Biulchi e Pauli (2012); Boroujerdi et al. (2016); Luz e Figueiredo Filho (2018); Dewettinck (2008); Dewettinck e van Dijk (2013); Haines e St-Onge (2012); Haines et al. (2004); Iqbal et al. (2019); Ishaq et al. (2009); Ishaq et al. (2013); Kakkar et al. (2020); Kang e Choi (2019); Kivipöld et al. (2020); Lawler (2003); Longenecker e Nykodým (1996); Longenecker et al. (1988); Lopes et al. (2020); Makhubela et al. (2016); Maley e Moeller (2014); Mok e Leong (2021); Monsur e Akkas (2015); Phuong (2018); Ponmuthumari (2020); Roberts (1992, 1995); Rolle e Klingner (2012); Sadiyawati e Hasanati (2024); Schleicher et al. (2019); Sharma et al. (2016, 2021); Soltani et al. (2003); Teeroovengadum et al. (2019); Upadhyay et al. (2020); Van Waeyenberg e Decramer (2018); Yang e Hsieh (2007); Yu et al. (2018)	40	68

Apesar dos critérios identificados como pertencentes à dimensão Transferência figurarem nos estudos revisados sobre efetividade de GD, pontua-se que isso se dá, principalmente, de maneira isolada ou pontual, sem serem relacionados a algum grupo específico, como uma dimensão. Logo, destaca-se que, assim como a dimensão de Aprendizagem enquanto resultado de GD, a dimensão Transferência também foi assim constituída recentemente, surgindo dessa forma no modelo de avaliação de efetividade de GD proposto pelos autores Schleicher et al. (2019). Para isso, os autores se inspiraram novamente na literatura de treinamento, porém, desta vez, para pontuar que a adoção da palavra “transferência”, em seu modelo de avaliação de efetividade de GD, deve ser compreendida em

seu sentido amplo, pois o intuito não é demonstrar como é feita a aplicação do que foi aprendido, tal como prevê a literatura de avaliação de treinamento, mas, sim, como os resultados multidimensionais do sistema de GD são expressos no trabalho.

Historicamente, os critérios aqui identificados como de Transferência são pouco estudados enquanto resultado de GD, tendo a sua aplicação aumentado somente nos últimos anos e mais em estudos empíricos; logo, a constituição dessa dimensão se faz importante por permitir compreender atitudes, comportamentos e resultados que são afetados pelo sistema de GD e que ultrapassam o contexto específico do sistema de GD ao serem “transferidos” para o trabalho (Schleicher et al., 2019). Segundo os autores, as dimensões descritas até aqui falavam muito mais sobre o contexto em si do sistema de GD, ou seja, como os envolvidos reagem perante o sistema de GD e o que eles diretamente adquirem ao participar da GD, mas ainda não eram capazes de demonstrar como esses resultados “retornavam” para o trabalho.

Em geral, os estudos revisados apresentaram implicações relativas a avaliar aspectos relacionados aos resultados dos indivíduos, como melhoria de desempenho, aumento na motivação e na satisfação no trabalho, comprometimento e engajamento no trabalho, promoção das relações interpessoais. Esses critérios foram recorrentemente adotados pelos estudos revisados de forma combinada entre eles mesmos da dimensão Transferência (*e.g.* atitudes no trabalho, como satisfação, que afetam motivação e desempenho e vice-versa), explicitadas adiante.

As relações estabelecidas entre os critérios da dimensão Transferência foram teorizadas a partir da teoria de troca social (*social exchange theory* – SET), a qual sugere que resultados positivos no trabalho são uma forma de retribuição do investimento feito pela organização, como o sistema de GD, em seu bem-estar pessoal e profissional, promovendo maior engajamento e reduzindo a intenção de sair da organização (Kakkar et al., 2020; Sharma et al., 2021). De forma semelhante, a teoria da equidade (*equity theory*) também fornece insumos importantes ao advogar que a percepção de equidade sobre o sistema de GD pode aumentar o engajamento do funcionário com a organização e seu trabalho e reduzir a sua intenção de deixar a organização (Sharma et al., 2021).

Essas duas teorias também foram usadas para demonstrar a importância das reações para compreender a efetividade do sistema de GD e adotadas novamente enquanto as bases teóricas utilizadas pelos estudos revisados para justificar as relações estabelecidas entre os critérios da dimensão Transferência e de outras dimensões estão vinculadas ao campo da Psicologia.

A dimensão Transferência, nos sistemas de GD, é composta pelos critérios de Desempenho no trabalho, Motivação, Atitudes no trabalho, Relação interpessoal e Bem-estar, respectivamente, conforme demonstrado na Tabela 8.

Tabela 8

Critérios da dimensão de Transferência

Critérios	Autores	Frequência	
		n	%
Desempenho no trabalho	Ahmed (1999); Awan et al. (2020); Luz e Figueiredo Filho (2018); Dewettinck (2008); Dewettinck e van Dijk (2013); Haines e St-Onge (2012); Haines et al. (2004); Ishaq et al. (2013); Kang e Choi (2019); Lawler (2003); Longenecker e Nykodym (1996); Longenecker et al. (1988); Lopes et al. (2020); Monsur e Akkas (2015); Phuong (2018); Ponmuthumari (2020); Roberts (1992); Roberts (1995); Sadiyawati e Hasanati (2024); Schleicher et al. (2019); Yu et al. (2018)	21	36
Motivação	Biulchi e Pauli (2012); Dewettinck (2008); Dewettinck e van Dijk (2013); Ishaq et al. (2009); Ishaq et al. (2013); Lawler (2003); Longenecker e Nykodym (1996); Longenecker et al. (1988); Maley e Moeller (2014); Mok e Leong (2021); Monsur e Akkas (2015); Roberts (1992); Roberts (1995); Schleicher et al. (2019); Soltani et al. (2003); Van Waeyenberg e Decramer (2018); Yang e Hsieh (2007); Yu et al. (2018);	18	32
Atitudes no trabalho	Abdullah e Abraham (2012); Awan et al. (2020); Boroujerdi et al. (2016); Dewettinck (2008); Dewettinck e van Dijk (2013); Govender e Bussin (2020); Iqbal et al. (2019); Kakkar et al. (2020); Lawler (2003); Makhubela et al. (2016); Mok e Leong (2021); Schleicher et al. (2019); Sharma et al. (2016); Sharma et al. (2021); Teeroovengadum et al. (2019); Upadhyay et al. (2020)	16	28
Relação Interpessoal	Boroujerdi et al. (2016); Kivipöld et al. (2020); Longenecker e Nykodym (1996); Longenecker et al. (1988); Maley e Moeller (2014); Mok e Leong (2021); Ponmuthumari (2020); Rolle e Klingner (2012); Schleicher et al. (2019); Upadhyay et al. (2020); Yu et al. (2018);	11	19
Bem-estar	Biulchi e Pauli (2012); Dewettinck e van Dijk (2013); Schleicher et al. (2019)	3	5

O critério mais adotado da dimensão Transferência foi Desempenho no trabalho, identificado em 21 estudos revisados (36%). Definir precisamente o construto “desempenho” não é fácil, como visto até aqui, e o conteúdo dos estudos revisados reforçam isso ao não apresentarem uma definição conceitual precisa. Assim, foi utilizada como definição o que diz Campbell (1983) referente ao nível em que o indivíduo ajuda a organização a alcançar seus resultados, conforme referenciaram Motowilde, Borman e Schmit (1997), os quais, em sua pesquisa utilizada como referência no estudo revisado dos autores Awan et al. (2020, p. 72), complementam que desempenho é “episódico, avaliável e multidimensional”. Em linhas gerais, pode-se dizer que desempenho no trabalho consiste no comportamento expresso no trabalho pelo indivíduo que auxilia no alcance de resultados organizacionais.

A maioria dos autores que adotaram esse critério para avaliar a efetividade do sistema de GD referiram-se como a capacidade do sistema de GD em promover melhorias no desempenho individual no trabalho de forma genérica, isto é, sem especificar quais elementos estavam sendo considerados (*e.g.* Luz & Figueiredo Filho, 2018; Dewettinck, 2008; Dewettinck & van Dijk, 2013; Haines & St-Onge, 2012; Haines et al., 2004; Ishaq et al., 2009; Kang & Choi, 2019; Lawler, 2003; Longenecker & Nykodym, 1996; Longenecker et al., 1988; Lopes et al., 2020; Monsur; Akkas, 2015; Phuong, 2018; Ponmuthumari, 2020; Roberts, 1995; Sadiyawati; Hasanati, 2024, Yu et al., 2018). Esse mesmo achado também foi apontado por Schleicher et al. (2019) em sua revisão.

Nesse sentido, os estudos utilizavam perguntas direcionadas aos funcionários por meio de questionário em relação a como o sistema de GD “tem claros efeitos sobre o meu desempenho” e “me faz funcionar melhor no trabalho” (*e.g.* Dewettinck; Van Dijk, 2013); outra pergunta em outro estudo revisado consistiu na percepção de funcionários da organização sobre como o sistema de GD “melhora o desempenho a partir do estabelecimento de indicadores de desempenho” (*e.g.* Kang & Choi, 2019). No estudo de Costigan et al. (2009), foi avaliado em que medida a efetividade do sistema de GD promove um comportamento confiável do funcionário, suficiente para representar uma contribuição em suas responsabilidades e outro comportamento avaliado foi honrar com os compromissos estabelecidos de trabalho. Segundo apontado por Roberts (1992), desempenhos inferiores são capazes de serem reduzidos, quando o sistema de GD é aceito pelos funcionários, desenvolve suas habilidades, motiva-os, promove aumento na produtividade e reduz o absenteísmo.

A classificação de desempenho em intrapapel (tarefa) e extrapapel (contextual) é adotada por Awan et al. (2020), referenciando Motowilde et al. (1997). Para facilitar a

compreensão e a mensuração do critério, o desempenho de tarefa consiste em mudanças no comportamento nas tarefas do trabalho diretamente relacionadas à consecução de objetivos ou metas organizacionais. Já o desempenho contextual é referente às ações que vão além das atribuições de trabalho formalmente estabelecidas. Neste estudo, as variáveis de desempenho de tarefa e contextual foram mensuradas por duas escalas distintas: para o desempenho de tarefa, foi adotada a escala de Van Dyne e Lepine (1988) e para o contextual, a escala de Borman e Motowildo (1993). Essa subclassificação do desempenho no trabalho também foi apontada no estudo de Schleicher et al. (2019) ao identificarem que apenas o desempenho de tarefa figurou em estudos teóricos de avaliação de efetividade de GD revisados por eles e apenas em alguns poucos exemplos.

O segundo critério mais adotado dessa dimensão foi Motivação, identificado em 18 artigos revisados (32%), os quais avaliaram a capacidade do sistema de GD em melhorar a disposição ou a tendência em exercer esforços no trabalho conforme procedimentos e políticas da organização (Jiang et al., 2013; Latham et al., 2005). Essa definição decorre das teorias de CO e da psicologia industrial e organizacional, campos de estudos adotados pelos estudos de avaliação de efetividade de GD. Assim como desempenho, a Motivação não teve uma definição conceitual precisa identificada para esse critério nos artigos revisados.

Novamente, assim como aconteceu em alguns critérios da dimensão de Aprendizagem, a maior parte dos estudos são genéricos em avaliar que o sistema de GD “motiva seus funcionários” no trabalho, sem indicar em que sentido ocorre essa motivação (*e.g.* Ishaq et al., 2009; Ishaq et al., 2013; Longenecker et al., 1988; Longenecker & Nykodym, 1996; Maley; Moeller, 2012; Monsur; Akkas, 2015; Roberts, 1992, 1995; Schleicher et al., 2019; Yang & Hsieh, 2007; Yu et al., 2018).

No estudo de Monsur e Akkas (2015), questiona-se aos funcionários de uma universidade de Bangladesh em que medida eles acreditam que o sistema de GD os mantém motivados. Os autores concluem que o sistema de GD é tido como capaz de promover efetividade desde que ele tenha sido adequadamente proposto e implementado na organização. Segundo Mok e Leong (2021), o sistema de GD está diretamente relacionado à motivação do funcionário, a qual é impactada pela percepção de que o sistema é justo e, a partir disso, estabelece-se relação também com a satisfação no trabalho e o compromisso organizacional. Para os autores, funcionários motivados auxiliam a alcançar objetivos organizacionais e reduzir taxas de rotatividade. Para os autores Biulchi e Pauli (2012), em seu estudo sobre o sistema de GD em uma organização pública de seguridade social no Brasil, a motivação é apontada por impactar positivamente na produtividade e, assim, entregar serviços de qualidade ao cidadão.

A efetividade do sistema de GD foi avaliada por Dewettinck (2008) a partir da criação de uma Escala do Efeito Motivacional para Revisão do Desempenho, com foco apenas na etapa de revisão do desempenho. A escala foi aplicada com gestores de GP a fim de identificar a capacidade do sistema de GD em aumentar o desempenho e a orientação da organização para resultados, em contraponto a promover motivação, desenvolvimento e colaboração do funcionário. O autor identificou que um sistema de GD é, na perspectiva dos gestores das unidades de GP, menos efetivo se ele é voltado para motivar funcionários do que quando o sistema de GD é desenhado para a promoção de desempenho superior no trabalho. Logo, percebe-se que, para os gestores de GP, é mais importante um sistema de GD cujo objetivo está voltado para a promoção de desempenho individual do que para motivar seus funcionários, por mais que se saiba que a motivação também pode promover desempenho a longo prazo. Nesse sentido, o autor conclui que os gestores de GP são mais propensos a avaliar a efetividade de um sistema de GD quando estes promovem resultados de curto prazo, uma vez que melhorias na motivação dos funcionários também poderiam acarretar melhorias no desempenho, só que em uma perspectiva de longo prazo.

Identificou-se, também, que alguns autores estabelecem uma relação entre motivação, feedback e reconhecimento. Lawler (2003) destaca que reconhecer sistematicamente é efetivo em promover motivação no trabalho. Dewettinck (2008) também destaca o papel central do feedback intermediário entre supervisores e empregados para promover a motivação dos funcionários. Soltani et al. (2003) também ressaltaram a importância de feedbacks direcionados para motivar os funcionários, evitando interpretações voltadas para autoimagens de sucesso ou de fracasso. Iqbal et al. (2013) afirma que a escassez de estudos sobre motivação do funcionário e feedback é a razão para sistemas de GD ineficazes.

O terceiro critério mais identificado foi Atitudes no trabalho, encontrado em 16 artigos revisados (28%). Essa categoria foi constituída diante dessa definição que contempla aspectos como: engajamento no trabalho, comprometimento organizacional e satisfação no trabalho. Assim, nenhum estudo revisado apresentou a definição específica para o critério de Atitudes no trabalho, tendo sido utilizada a definição de Judge e Kammeyer-Muller (2012), que expressam os sentimentos, crenças e apego ao trabalho. Segundo os autores, esses aspectos compõem a categoria chamada de “*job attitudes*”, em inglês. Esse mesmo critério foi identificado dessa forma no estudo de Schleicher et al. (2019), porém, como dito, sem uma definição constitutiva.

Os estudos avaliaram que a efetividade do sistema de GD pode ser aferida a partir da promoção de comprometimento organizacional (e.g. Iqbal et al., 2019; Lawler, 2003). O comprometimento organizacional é referente ao “vínculo afetivo com a organização, um

sentimento de lealdade em relação a ela e a intenção de permanecer como parte dela” (Judge; Kammeyer-Muller, 2012, p. 343). Outros aspectos avaliados foram: 1) engajamento no trabalho, como sendo resultado da combinação de motivações intrínsecas (atender aos objetivos pessoais e “encher” o ego) e extrínsecas (desejo de obter recompensas e evitar punições) (*e.g.* Awan et al., 2020; Kakkar et al., 2020; Lopes et al., 2020; Sharma et al., 2021), sendo que, segundo Awan et al. (2020), feedback adequado é capaz de promover engajamento no trabalho; 2) colaboração (*e.g.* Dewettinck, 2008; Dewettinck & van Dijk, 2013); 3) envolvimento (*e.g.* Teerovengadum et al., 2019); e 4) participação no trabalho (*e.g.* Boroujerdi et al., 2016). Os autores também apontaram enquanto resultado a satisfação com o trabalho (*e.g.* Kakkar; Mok & Leong, 2021; Schleicher et al., 2019; Upadyay et al., 2020) em que são expressos sentimentos positivos e contentamento com o trabalho (Judge et al., 2012).

A relação entre satisfação com o trabalho e intenções de saída da organização enquanto critérios de efetividade do sistema de GD pode ser compreendida a partir da aplicação da teoria da troca social (*SET*), a qual permite entender que, a partir da relação de troca social, resultados individuais positivos (como satisfação e comprometimento com o trabalho) são vistos como trocas recíprocas com os membros da organização, práticas e sistemas, permitindo compreender que o sistema de GD é capaz de promover crescimento profissional e bem-estar na organização (Kakkar et al., 2020).

O quarto critério identificado, também com 11 estudos revisados (19%) que adotaram esse critério, foi Relação Interpessoal, referindo-se, em geral, às mudanças na interação social entre os indivíduos da organização (Schleicher et al., 2019). Os estudos identificaram que a efetividade de um sistema de GD pode ser avaliada pela melhora nas relações interpessoais, principalmente entre gestor e subordinado (*e.g.* Longenecker & Nykodym, 1996; Longenecker et al., 1988; Kivipöld et al., 2020; Ponmuthumari, 2020; Upadhyay et al., 2020; Yu et al., 2018). Maley e Moeller (2014), em seu estudo sobre a efetividade do sistema de GD em multinacionais de diferentes países, buscaram avaliar a partir da perspectiva das correspondências remotas entre gestores e funcionários e identificaram que eles percebem que o sucesso do sistema de GD está atrelado à frequência e à natureza das interações supervisor-supervisionado, e estas devem ser pautadas, segundo eles, em uma “relação construtiva”, sem mencionar quais aspectos deveriam ser considerados.

Ainda com relação às trocas entre gestores e funcionários, Kivipöld et al. (2020) e Ponmuthumari (2020) avaliaram se o sistema de GD foi capaz de favorecer um espaço de abertura para discussão ou sugestões entre esses dois grupos de indivíduos. Por outro lado, no estudo de Rolle e Klingner (2012), os funcionários avaliaram que seus supervisores não

possuem as habilidades necessárias para efetivamente implementar um sistema de GD e que isso influencia na efetividade do sistema, gerando sentimentos de desconfiança e incerteza entre os funcionários.

Outro aspecto avaliado foi o nível de comunicação (Boroujerdi et al., 2016; Mok & Leong, 2021). Isso implica que as informações repassadas sejam claras e compreendidas pelos indivíduos que recebem a mensagem. Entretanto, no estudo de Mok e Leong (2021), os autores avaliaram que o nível de comunicação não tem um impacto significativo na efetividade de GD.

Por último, o critério Bem-estar, referente ao aumento do prazer e a redução do sofrimento com o trabalho (Schleicher et al., 2019), foi o menos adotado dessa dimensão, identificado em três artigos, correspondendo a 5% da amostra de estudos revisados (*e.g.* Biulchi; Pauli, 2012; Dewettinck & van Dijk, 2013; Schleicher et al., 2019).

Ao avaliarem a efetividade do sistema de GD, Biulchi e Pauli (2012) identificaram que um possível impacto no salário dos funcionários decorrente das ações do sistema de GD exerce influência na segurança monetária dos trabalhadores, alterando o sentimento de felicidade ou sofrimento no trabalho. Por último, Dewettinck e van Dijk (2013) avaliaram se um sistema de GD efetivo era capaz de favorecer a autoestima do funcionário.

Portanto, essa dimensão aborda os resultados do sistema de GD que são transferidos para o trabalho e que foram identificados como sendo relacionados ao desempenho, motivação, atitudes no trabalho, relação interpessoal e bem-estar. Percebeu-se que, em geral, os estudos não especificam o que está sendo considerado em cada um dos critérios identificados, deixando de uma maneira abrangente e genérica.

Com relação às características da coleta dos estudos que adotaram a dimensão Transferência, destaca-se que apenas um estudo utilizou dados objetivos e não perceptuais de nível organizacional, como um indicador (*proxy*) de transferência, ao avaliar um dos resultados da efetividade de GD relativo à Transferência. É o caso dos autores Biulchi e Pauli (2012), que utilizaram o indicador organizacional de “Idade Média do Acervo”, referente ao tempo de espera para atendimento em uma agência de seguridade social no Brasil, para precisar que quanto menor a quantidade de dias de represamento, mais motivados os servidores da agência se mostraram, segundo o estudo.

Assim como as duas dimensões anteriores, esta também é predominantemente quantitativa (85%) com dados primários (77%), coletados por meio de questionários (67%) (*e.g.* Ahmed, 1999; Awan et al., 2020; Dewettinck, 2008; Dewettinck & van Dijk, 2013; Haines & St-Onge, 2012; Haines et al., 2004; Iqbal et al., 2019; Ishaq et al., 2009; Ishaq et al., 2013; Kakkar et al., 2020; Kang & Choi, 2019; Lawler, 2003; Longenecker & Nykodym, 1996;

Makhubela et al., 2016; Mok & Leong, 2021; Phuong, 2018; Ponmuthumari, 2020; Roberts, 1992; Roberts, 1995; Sharma et al., 2021; Soltani et al., 2003; Teeroovengadum et al., 2019; Upadhyay et al., 2020; Van Waeyenber; Decramer, 2018; Yang & Hsieh, 2007; Yu et al., 2018). Esses questionários foram aplicados no nível individual de forma subjetiva, feitos aos múltiplos perfis de participantes do sistema de GD (57%), o que inclui funcionários, diretos, gestores de GP e de linha, e da alta administração.

Na Transferência, apesar de ter o critério Desempenho no trabalho como o mais adotado, seguido pelo critério de Motivação (segundo mais adotado), não foi identificada a aplicação de nenhuma escala própria para sua avaliação nos estudos revisados, apenas figurando como itens nos instrumentos de avaliação de efetividade de GD. Assim, destaca-se o critério de Atitudes no trabalho, cuja medida com mais escalas foi a de percepção de satisfação no trabalho, como a de *Michigan Organizational Assessment Questionnaire*, de Bowling e Hammond (2008), adotado por Kakkar et al. (2020), Escala de Satisfação no Trabalho, de Macdonald e MacIntyre (1997) (Sharma et al., 2016), e a Escala de Satisfação em Geral, de Taylor e Bowers (1972) (Upadhyay et al., 2020); outra escala aplicada foi *Utrecht Work Engagement Scale (UWES)*, de Schaufeli et al. (2006), no caso da medida de engajamento no trabalho, a qual foi adotada por dois estudos revisados: Kakkar et al. (2020) e Sharma et al. (2021); já a percepção de comprometimento no trabalho teve duas escalas identificadas: de Angle e Perry, em Sharma et al. (2016) e de Kalleberg, Knoke, Marsden, e Spaeth (1996) (Iqbal et al., 2019).

2.3.1.4 Dimensão de Resultados Operacionais

Com relação aos Resultados Operacionais, estes constituem a quarta dimensão mais adotada e referem-se às mudanças nos aspectos relacionados à operação de processos organizacionais (Schleicher et al., 2019), tendo essa dimensão sido adotada em 16 estudos, correspondendo a 28% da amostra revisada, conforme disposto na Tabela 9.

Tabela 9

Estudos revisados que adotaram a dimensão Resultados Operacionais

Dimensão	Estudos revisados	Frequência	
		n	%

Resultados Operacionais	Biulchi e Pauli (2012); Espejo et al. (2021); Haines e St-Onge (2012); Haines et al. (2004); Ishaq et al. (2009); Kang e Choi (2019); Longenecker e Nykodym (1996); Longenecker et al. (1988); Lopes et al. (2020); Nankervis et al. (2012); Roberts (1992, 1995); Schleicher et al. (2019); Stanton e Nankervis (2011); Upadhyay et al. (2020); Yang e Hsieh (2007)	16	28
--------------------------------	--	----	----

Enquanto as dimensões citadas anteriormente tinham como foco avaliar resultados mais imediatos, ou seja, aqueles mais próximos ao sistema de GD e que impactam mais diretamente os indivíduos, a dimensão Resultados Operacionais é a primeira categoria proposta a considerar os resultados cujo fenômeno acontece à nível unidade/organizacional, e não mais individual. Isso significa que os critérios pertencentes a essa dimensão podem ser tanto relacionados a informações da unidade de GP ou outra, como também aplicados a dados de toda a organização.

É importante pontuar que os resultados do sistema de GD mais distais são esses relacionados ao desempenho organizacional em que os resultados se distanciam do nível individual e se aproximam daqueles de macronível, cujos esforços individuais são agregados (Dyer & Reeves, 1995). Ressalta-se que considerar critérios pertencentes ao nível unidade/organizacional é compreensível, especialmente porque demonstrar o impacto de modelos de GP no desempenho organizacional é um desafio. Nesse contexto, torna-se mais viável e esclarecedor, tanto para pesquisadores quanto para gerentes, buscar evidências do impacto das práticas e políticas de GD em resultados mais operacionais ou intermediários, que, por sua vez, podem estar relacionados a resultados mais distantes, como desempenho financeiro ou de mercado (Colakoglu et al., 2006). Assim, foram identificados os seguintes critérios: Rotatividade, Produtividade, Inovação e Força de trabalho, respectivamente, em relação à quantidade de estudos que os adotaram, conforme demonstrado na Tabela 10.

Tabela 10

Crítérios da dimensão Resultados Operacionais

Crítérios	Autores	Frequência	
		n	%
Produtividade	Biulchi e Pauli (2012); Kang e Choi (2019); Longenecker e Nykodym (1996); Longenecker et al. (1988); Roberts (1992); Roberts (1995); Schleicher et al. (2019); Staton e Nankervis (2011); Upadhyay et al. (2020); Yang e Hsieh (2007)	10	18

Rotatividade	Espejo et al. (2021); Haines e St-Onge (2012); Haines et al. (2004); Ishaq et al. (2009); Roberts (1992); Roberts (1995); Schleicher et al. (2019)	7	12
Inovação	Haines e St-Onge (2012); Haines et al. (2004); Lopes et al. (2020); Nankervis et al. (2012); Schleicher et al. (2019); Staton e Nankervis (2011)	6	11
Força de trabalho	Nankervis et al. (2012); Schleicher et al. (2019); Staton e Nankervis (2011)	3	5

O primeiro critério mais adotado dessa dimensão foi Produtividade, com dez estudos (18%), os quais se referem à capacidade de aumentar a quantidade ou a qualidade do trabalho realizado com uma quantidade igual ou menor de recursos (Schleicher et al., 2019). Em geral, os estudos indicam que a efetividade de GD pode ser aferida a partir do aumento de produtividade, sem especificar quais aspectos poderiam ou foram considerados (*e.g.* Kang & Choi, 2019; Longenecker & Nykodym, 1996; Longenecker et al., 1988; Schleicher et al., 2019; Roberts, 1995; Schleicher et al., 2019; Staton; Nankervis, 2011; Upadhyay et al., 2020; Yang & Hsieh, 2007; Yu et al., 2018). Como, por exemplo, ao indagar a funcionários, por meio de questionários, em que medida eles acreditam que o sistema de GD foi/é capaz de promover produtividade (*e.g.* Kang & Choi, 2019; Longenecker & Nykodym, 1996; Longenecker et al., 1988; Roberts, 1995).

O único estudo a utilizar dados objetivos (*hard*) a partir da coleta de documentos organizacionais foi o de Biulchi e Pauli (2012) que identificaram que a quantidade de dias de represamento da demanda (Idade Média do Acervo) foi reduzida após o sistema de GD em uma organização pública. O estudo de Nankervis et al. (2012) teve como objetivo avaliar a relação do sistema de GD com resultados organizacionais a partir da perspectiva de diferentes níveis de gestores de organizações e indústrias da Austrália, os quais não identificaram diferença na percepção da melhora na qualidade dos serviços prestados após o sistema de GD.

O segundo critério mais adotado foi Rotatividade, identificado em sete estudos revisados (12%), e se refere à quantidade de admissões, demissões e absenteísmo do funcionário na organização (Schleicher et al., 2019). Nesse sentido, a redução no absenteísmo (*e.g.* Roberts, 1992; 1995), promoção da retenção de funcionários (*e.g.* Haines et al., 2004; Haines & St-Onge, 2012; Roberts, 1995) e a redução na taxa de rotatividade (*e.g.* Ishaq et al., 2009) foram aspectos sugeridos pelos modelos teóricos de avaliação de efetividade do sistema de GD ou testados empiricamente a partir da percepção de melhoria na relação de admissões e demissões a partir do sistema de GD.

Na pesquisa de Espejo et al. (2021), avaliou-se a efetividade do sistema de GD em uma concessionária elétrica no Brasil e identificou-se que não há relação entre um sistema de GD efetivo e a taxa de demissões, pois as notas consideradas boas ou excelentes dos avaliados não impediram que eles fossem posteriormente desligados. Nesse sentido, isso implica dizer que, tendo em vista o objetivo do sistema de GD de gratificar os funcionários com desempenho elevado, a implementação da GD pode não se mostrar vantajosa para a organização, tendo em vista que aqueles com resultados superiores na avaliação de desempenho ainda assim foram em seguida desligados da organização, o que pode representar perda de investimento para a organização.

O terceiro critério mais adotado da dimensão de Resultados Operacionais foi a Inovação, utilizada em seis estudos (11%), referente ao desenvolvimento de novas ferramentas ou mudanças no processo de trabalho decorrentes do sistema de GD (Schleicher et al., 2019). A identificação desse critério nos estudos se restringe à indicação de que a efetividade do sistema de GD pode ser avaliada a partir de melhorias nos processos organizacionais, porém sem relatar como ou em que aspectos essa inovação deveria se dar (*e.g.* Haines & St-Onge, 2012; Haines et al., 2004; Nankervis et al., 2012; Schleicher et al., 2019; Stanton; Nankervis, 2011).

Esse critério também foi avaliado no estudo de Lopes et al. (2020) em que os servidores agentes implementadores do programa de desenvolvimento que inclui o sistema de GD em uma organização pública não visualizaram mudanças significativas na execução dos processos de trabalho, considerando a implantação da AD. Segundo eles, a inovação trazida foi a consideração dos planos de trabalho (onde se impõe atividades, prazos etc.) como um meio para a efetividade de GD, entretanto os dados revelaram que isso não representou melhorias no *modus operandi* com que o programa é conduzido na organização.

Por último, o critério Força de Trabalho foi adotado em quatro estudos (7%) e referiu-se à melhora na qualidade e/ou quantidade do quadro de funcionários da organização (Schleicher et al., 2019). Para Stanton e Nankervis (2011), a importância de o sistema de GD ocasionar melhorias na qualidade da força de trabalho foi citada pelos participantes gestores de GP do estudo de organizações públicas e privadas ao serem questionados sobre os resultados do sistema de GD, os quais reconheceram a sua importância enquanto investimento para efetividade e competitividade organizacional. Nankervis et al. (2012) e Stanton e Nankervis (2011) também identificaram outros aspectos de nível organizacional como financeiros (*e.g.* lucro e participação no mercado), citados pelos participantes de forma conjunta com os resultados operacionais (*e.g.* força de trabalho, produtividade e inovação).

Em resumo, a dimensão de Resultados Operacionais é a primeira a fornecer informações importantes sobre os resultados do sistema de GD a nível de unidade/organização, como também identificado no estudo de Schleicher et al. (2019). Porém percebeu-se, ainda, uma incipiência nos resultados encontrados em razão da quantidade relativamente pequena de estudos revisados, o que dificulta uma discussão mais aprofundada sobre os critérios de avaliação de efetividade de GD, mesmo diante da importância relatada pelos próprios estudos, que atribuem ao sistema de GD a capacidade de promover resultados organizacionais. Nesse sentido, esse é o reflexo do que fora apontado por Ikramullah et al. (2016) sobre quando a literatura de um campo prioriza determinadas dimensões e desconsidera outras, subvertendo conteúdos que poderiam promover novas discussões sobre o tema de interesse para outros atores organizacionais (*e.g.* acionista, diretor etc.).

Alterando o nível do fenômeno de individual para organizacional, percebem-se também alterações nas características metodológicas dos estudos que adotaram essa dimensão de Resultados Operacionais. Nesse sentido, a amostra continua sendo predominantemente quantitativa (81%), sendo que nenhum estudo foi somente qualitativo; o uso foi majoritário de dados primários (81%), porém destaca-se que essa dimensão possui, proporcionalmente, mais estudos com dados secundários (19%).

Com relação ao instrumento de coleta, assim como a dimensões anteriores, também se mostrou predominante o uso de questionários (69%), utilizado para capturar a percepção de múltiplos perfis da organização (75%), o que inclui funcionários, gestores de GP, de linha e alta administração; destaca-se a quantidade proporcional de estudos que utilizaram pesquisa documental (13%), que aumentou em relação às demais dimensões em razão da adoção de medidas objetivas para avaliar os critérios de Produtividade (*e.g.* Biulchi; Pauli, 2012) e Rotatividade (*e.g.* Espejo et al., 2021).

Destacam-se os critérios de Produtividade e de Rotatividade como os únicos a utilizarem medida objetiva de nível organizacional, como a quantidade de tempo médio de espera de atendimento (Idade Média do Acervo) no estudo de Biulchi e Pauli (2012) enquanto medida de produtividade e a taxa de desligamento em Espejo et al. (2021) para aferir a rotatividade. A utilização desse tipo de dado inova até aqui ao adotar dados objetivos em meio à adoção predominante de dados subjetivos, inclusive em critérios passíveis de mensuração por meio de dados objetivos. Logo, a adoção dos critérios de produtividade e rotatividade com dados objetivos pode ampliar a discussão sobre a efetividade de GD, conforme afirma Wall et al. (2004). Portanto, a dimensão de Resultados Operacionais tem se diversificado

metodologicamente, mesmo que ainda de maneira limitada e reduzida, das demais dimensões anteriores.

2.3.1.5 Dimensão de Resultados Financeiros

A penúltima e menos adotada dimensão foi a de Resultados Financeiros, que consiste nos aspectos relacionados a orçamento, ações, produtos e serviços financeiros, econômicos ou contábeis (Nankervis et al., 2012). Essa dimensão foi adotada em três estudos, correspondendo a 5%, conforme contido na Tabela 11.

Tabela 11

Estudos revisados que adotaram a dimensão Resultados Financeiros

Dimensão	Estudos revisados	Frequência	
		n	%
Resultados Financeiros	Nankervis et al. (2012); Schleicher et al. (2019); Stanton e Nankervis (2011)	3	5

Os Resultados Financeiros, assim como os operacionais, também buscam avaliar os resultados do sistema de GD a nível unidade/organizacional. Por essa razão, o desempenho financeiro é apontado pela literatura de GP como a última “razão de ser” de um sistema em uma organização (Becker; Huselid, 1998), sendo recorrentemente adotado em estudos que avaliam os efeitos de práticas estratégicas de GP (*e.g.* Huselid, 1995; Cascio, 1991; Jiang et al., 2012). Essa dimensão figura como um dos resultados de sistema de GP defensáveis, segundo apontado por Dyer e Reeves (1995), além dos operacionais e daqueles relacionados à GP. Logo, como forma de mensurar resultados de um sistema estratégico de pessoal, é recorrente encontrar menção a esse estudo na literatura de GP para justificar a sua adoção (*e.g.* Becker; Huselid, 2006; Colakoglu et al., 2006; Paauwe & Boselie, 2005).

Entretanto, no caso da literatura de GD, essa dimensão ainda é pouco trabalhada, conforme é possível inferir pela frequência de estudos que avaliaram os resultados do sistema de GD a partir dos resultados financeiros, os quais também não realizaram um estudo aprofundado dessa dimensão. A sua presença reduzida na avaliação de resultados do sistema de GD também foi identificada na revisão realizada por Schleicher et al. (2019). Logo, esta revisão

não tem como intenção ser taxativa nos conceitos financeiros das organizações, mas demonstrar os critérios adotados pelos estudos na avaliação dos resultados do sistema de GD.

Foram identificados três critérios: Lucratividade, Participação no mercado e Retorno financeiro, conforme Tabela 12.

Tabela 12

Critérios da dimensão Resultados Financeiros

Critérios	Autores	Frequência	
		n	%
Lucratividade	Nankervis et al. (2012); Schleicher et al. (2019); Staton e Nankervis (2011)	3	5
Retorno em investimento	Nankervis et al. (2012); Schleicher et al. (2019); Staton e Nankervis (2011)	3	5
Participação no mercado	Nankervis et al. (2012); Schleicher et al. (2019); Staton e Nankervis (2011)	3	5

Os três critérios possuem a mesma frequência ($n = 3$, 5%), sendo Lucratividade o indicador do ganho financeiro obtido, retirando-se tributos, custos e despesas, da organização. Esse critério foi adotado pelos três estudos: Nankervis et al. (2012), Schleicher et al. (2019) e Stanton e Nankervis (2011). Os critérios Retorno financeiro, o qual indica o ganho financeiro após um investimento, e Participação no mercado, referente à parcela de determinado produto ou serviço que a organização detém no mercado, foram considerados pelos três mesmos estudos, correspondendo a 5% da amostra dos estudos revisados: Nankervis et al. (2012), Schleicher et al. (2019) e Stanton e Nankervis (2011). No caso do Retorno financeiro, Nankervis et al. (2012) e Stanton e Nankervis (2011) consideraram como variável Retorno sobre investimento, referente a quanto volta à organização a partir do investimento feito. Essa mesma métrica também foi identificada na revisão de Schleicher et al. (2019). Já no caso da Participação de mercado, o qual indica a quantidade do mercado em que a organização está inserida e que ela domina, foi avaliado nos estudos de Nankervis et al. (2012) e Stanton e Nankervis (2011) e identificado em Schleicher et al. (2019) como competitividade do mercado.

Por fim, percebe-se que os estudos estabeleceram poucas relações das dimensões anteriores com os Resultados Financeiros e destes com as demais dimensões identificadas. Isso se deve, em partes, pela visão em conjunto por parte de alguns estudos revisados de que os Resultados Operacionais e Resultados Financeiros constituem “Resultados Organizacionais” (e.g. Nankervis et al., 2012), logo, as relações estabelecidas por esses autores em relação aos resultados operacionais podem também ser interpretadas como parte dos resultados financeiros.

Com relação às características metodológicas, dois dos três estudos que adotaram essa dimensão são mistos (75%), ou seja, qualitativa e quantitativa, sendo um estudo de Nankervis et al. (2012), em que foi utilizada uma pergunta aberta no questionário aplicado, e o outro sendo a revisão sistemática de Schleicher et al. (2019); não houve nenhum estudo somente qualitativo. Acerca da origem dos dados, dois estudos (50%) utilizaram dados primários e secundários. Segundo Lepak et al. (2006) e Colakoglu et al. (2006), como as organizações têm cada vez mais tido que reportar e registrar dados organizacionais, os estudos devem utilizar esses dados secundários como evidências para os resultados. Para a coleta dos dados, foi aplicado questionário em dois estudos (67%). O questionário foi aplicado apenas com gestores no estudo de Nankervis et al. (2012) e múltiplos perfis em Staton e Nankervis (2011).

O que mais chama a atenção na operacionalização dessa dimensão por parte dos estudos é a análise feita a nível individual e com medidas de natureza subjetiva, ou seja, a partir da percepção dos indivíduos participantes das pesquisas. Portanto, novamente se repete a predominância de medidas individuais para mensuração de fenômenos de nível macro-organizacional, conforme apontado por teóricos sobre o processo de psicologização (*e.g.* Budd, 2020; Godard, 2014; Kaufman, 2020). Além disso, os estudos continuam utilizando medidas de natureza subjetiva, apesar de autores que estudam a avaliação de resultados (*e.g.* Boselie et al., 2005; Wall et al., 2004) afirmarem que estudos que adotam resultados financeiros da organização devem buscar adotar resultados objetivos no lugar de resultados subjetivos. Esse comportamento também é identificado na literatura de avaliação de resultados no estudo não revisado de Wall et al. (2004), mas que fornecem importantes *insights* sobre a relevância das medidas de natureza de medidas objetivas, principalmente em relação a resultados e indicadores financeiros.

2.3.1.6 Dimensão de Impacto Societal

A última dimensão identificada é Impacto Societal, com quatro estudos (7%), conforme demonstrado na Tabela 13, a qual se refere às contribuições do que é planejado e entregue pela organização ao seu mundo externo, incluindo sociedade, clientes, fornecedores, entre outros *stakeholders* (Ikramullah et al., 2016).

Tabela 13

Estudos revisados que adotaram a dimensão Impacto Societal

Dimensão	Estudos revisados	Frequência	
		n	%
Impacto Societal	Biulchi e Pauli (2012); Ikramullah et al. (2016); Pinto e Behr (2015); Yang e Hsieh (2007)	4	7

A dimensão Impacto Societal identificada nos estudos revisados fornece informações quanto aos resultados obtidos no último nível de avaliação e, logo, o mais distante de todos do sistema de GD. Ao mencionar esse último nível, é importante delimitar que essa dimensão extrapola o nível organizacional e, por isso, é mencionada na literatura de avaliação em termos de “mega nível organizacional”, tratando daquilo que é planejado e entregue ao cliente e beneficiário primário (*e.g.* consumidor ou cidadão) (R. Kaufman & Keller, 1994).

Essa dimensão busca preencher uma lacuna de pesquisa de GEP de incluir aspectos do ambiente externo (Kaufman, 2020; Jiang et al., 2014), além de incluir outro grupo de *stakeholders* que não aqueles internos à organização (Colakoglu et al., 2006). O termo “impacto” aparece em alguns outros modelos de avaliação, porém como menção ao impacto no trabalho, a como alguns autores da literatura de avaliação de resultados de treinamento e de GP costumam se referir, ou simplesmente como uma menção genérica para relatar que os resultados do sistema de GD promovem impacto na organização. Portanto, não se deve confundir a dimensão Impacto Societal identificada neste estudo com as citadas anteriormente.

Já com relação aos estudos de avaliação do sistema de GD, essa dimensão não foi proposta na maior parte dos estudos empíricos revisados. Os únicos quatro estudos revisados que adotaram a dimensão Impacto Societal avaliaram o sistema de GD a partir da Satisfação do público-alvo, conforme demonstrado na Tabela 14.

Tabela 14

Crítérios da dimensão Impacto Societal

Crítérios	Autores	Frequência	
		n	%
Satisfação do público-alvo	Biulchi e Pauli (2012); Ikramullah et al., 2016; Pinto e Behr (2015); Yang e Hsieh (2007)	4	7

O critério de Satisfação do público-alvo, com quatro estudos (7%), considera o nível de contentamento do público-alvo com a organização em decorrência ou após a implementação do

sistema de GD na organização. A maioria dos estudos considerou a perspectiva dos cidadãos enquanto público-alvo da organização em razão de os estudos terem sido realizados em organizações públicas (e.g. Biulchi; Pauli, 2012; Pinto; Behr, 2015; Yang & Hsieh, 2007).

O estudo de Biulchi e Pauli (2012), para avaliação de efetividade do sistema de GD, identificou a redução do tempo de espera no atendimento do serviço público, o que concedeu segurança à população quanto à realização dos serviços de qualidade na região onde o estudo foi realizado. Já Pinto e Behr (2015) relatam que a avaliação de desempenho é vista como condição de melhoria para o serviço público. Por último, Yang e Hsieh (2007) avaliaram os resultados do sistema de GD no governo ao questionarem a percepção dos servidores públicos sobre a participação dos *stakeholders*, como cidadãos, na definição das estratégias de desempenho organizacional que guiam a GD na organização.

Para avaliar essa dimensão, cada um dos estudos empregou uma abordagem de pesquisa distinta: uma quantitativa (25%), outra qualitativa (25%) e uma última mista (25%), sendo um teórico, portanto, a abordagem não se aplica. Somente um estudo adotou dados secundários (25%) e outros dois fizeram uso de coleta de dados primários (50%). Os dados foram coletados tanto por meio de pesquisa documental (25%), entrevista com servidores técnico-administrativos de uma universidade (25) e questionário com servidores públicos de uma organização pública de Taiwan. Destaca-se que somente um estudo realizou o estudo no nível de análise organizacional e com medida objetiva para inferir a relação entre o sistema de GD e resultados de satisfação dos cidadãos; os outros dois realizaram no nível individual e com medidas de natureza subjetiva para indicar a satisfação do público-alvo.

Chama a atenção o impacto societal do sistema de GD ser desconsiderado pela grande parte dos estudos revisados, uma vez que uma avaliação de efetividade vai além das mudanças a nível individual, chegando até o nível de promoção de melhorias no nível ambiental em que a organização está inserida. Tal desconsideração pode estar atrelada à distância entre essa dimensão e a prática de GD, o que torna sua avaliação arriscada, sem a aplicação de técnicas estatística rigorosas que consigam estabelecer relações causais.

Por fim, o objetivo desta subseção consistiu em desenvolver um modelo teórico-integrativo a partir de revisão de literatura realizada, a qual permitiu identificar as dimensões e os critérios de avaliação de efetividade de GD adotados pelos estudos, traçando as relações apontadas pelos autores. A revisão também permitiu verificar os padrões metodológicos adotados pelos estudos para avaliar a efetividade de GD, o que resultou na identificação da adoção de medidas de nível individual.

Diante desta revisão sistemática de literatura, na próxima seção são discutidas as relações entre cada dimensão identificada. Conforme identificado, a literatura ainda está assentada na necessidade de confirmação de que o sistema de GD realmente interfere nos fenômenos de nível individual (reação, aprendizagem e transferência), ou seja, atuando como uma ferramenta capaz de modificar comportamentos individuais para que, a partir disso, possa promover desempenho organizacional superior. A identificação dessas relações será exposta adiante.

2.4 Relações Empíricas de Avaliação de Efetividade do Sistema de Gestão de Desempenho

A revisão de literatura empreendida buscou, primeiro, quantificar a frequência com que cada dimensão ou critério foi adotado (Tabela 2) e, a partir disso, traçar um panorama sobre de que forma os estudos revisados fazem escolhas por avaliar determinadas dimensões e critérios em detrimento de outros, identificando também os resultados obtidos e as técnicas e procedimentos metodológicos aplicados.

Diante disso, pontua-se que uma análise exclusiva dos resultados quantitativos da revisão sistemática de literatura realizada não é suficiente para responder a anseios de acadêmicos e praticantes. O fato de a dimensão Reação ter sido a mais frequente entre os estudos revisados não significa que ao avaliá-la exclusivamente será possível obter a efetividade de um sistema de GD em sua totalidade; assim como não é possível assumir que a dimensão menos adotada, de Impacto Societal, não forneça respostas sobre a efetividade de GD.

O que os resultados em conjunto da revisão trouxeram foram insumos que permitiram aferir quais são as dimensões e critérios avaliados e, destes, aqueles que mais são utilizados, além de identificar como essas dimensões e critérios se relacionam entre si. A construção e a validação de questionários podem representar importantes procedimentos a serem realizados por estudos futuros, a fim de identificar aquelas medidas que de fato devem integrar a dimensão e quais devem ser descartadas. A partir disso, um teste das relações entre as dimensões pode permitir identificar o quanto cada fator afeta o outro, chegando a um modelo mais enxuto de avaliação que, de certa forma, promova novas respostas aos anseios descritos anteriormente.

Nesse intento, torna-se relevante destacar que a revisão realizada também teve como resultado uma análise das relações e testes empíricos realizados pelos estudos a fim de avaliar a efetividade de GD. Identificou-se que a maior parte dos estudos consideram dimensões ou critérios cujo fenômeno é de nível individual ou de unidade/organizacional, internos ao

ambiente da organização (*e.g.* características da força de trabalho, rotatividade etc.). Esse achado mostra-se preocupante em razão da necessidade apontada de uma visão integrada e holística em que tanto aspectos internos quanto externos devem ser considerados, a fim de avaliar as diferentes formas com que a GD pode influenciar o desempenho organizacional (Cooke et al., 2021; Ikramullah et al., 2016).

A adoção de dimensões e critérios internos à organização guarda relação com a ordem lógica com que as dimensões de avaliação de efetividade de GD foram identificadas e se posicionam em relação ao sistema de GD. Isso porque as dimensões foram dispostas daquelas mais imediatas e próximas ao sistema de GD para aquelas menos e mais distantes, tal como apresentado nos modelos de avaliação de efetividade de GD (*e.g.* Levy & Williams, 2004b; Schleicher et al., 2019) e por aqueles modelos que buscam mensurar a efetividade de políticas e práticas de GP (*e.g.* Dyer & Reeves, 1995; Kaufman, 2020; Paauwe, 2009).

Estabelecer tais relações, buscando demonstrar causalidades de forma estatisticamente representativas entre o sistema de GD e as dimensões, principalmente aquelas mais distantes, como a de Impacto Social, mostra-se uma ação complexa e de difícil execução. Essa dificuldade obriga, de certa forma, que estudos realizem avaliações com as dimensões mais próximas ao sistema de GD e tidas como mais acessíveis (*e.g.* Ikramullah et al., 2016; Schleicher et al., 2019), pois estas aumentam as medidas sob controle da unidade/organização.

Dessa forma, a seguir são descritas as relações entre as variáveis do modelo que foram relatadas ou testadas na literatura revisada. O intuito não foi aprofundar exhaustivamente a análise sobre o quanto cada uma dessas variáveis impacta nas demais, mas, sim, identificar a existência de correlações e a direção dessas relações. Essa abordagem inicial permite mapear os principais caminhos e interações entre as variáveis.

2.4.1 Reação → Aprendizagem

A dimensão identificada como a mais imediata, ou seja, a primeira a ser considerada após o sistema de GD, é a de Reação. A sua predominância se assenta, primeiro, na sua definição e na sua própria natureza ao promover aos autores o entendimento de que esta é a forma com que os participantes do sistema de GD mais conseguem enxergar os benefícios do sistema (Keeping & Levy, 2000; Levy & Williams, 2004); e, segundo, na conveniência de acesso a esse tipo de informação por parte das unidades de GP a partir de levantamentos da percepção de tais medidas diretamente com os participantes de GD, tendo em vista se tratar de um fenômeno de nível individual, o qual pode ser facilmente obtido por meio de medidas

também de nível individual. Assim, destaca-se novamente que essa dimensão é composta por cinco critérios: Justiça, Precisão, Satisfação, Utilidade e Afetividade.

Com relação às relações estabelecidas entre a dimensão Reação e as demais, Keeping e Levy (2000) relacionam o fato de as reações estarem teoricamente relacionadas à aceitação (Atitude/Aprendizagem) do sistema de GD. Complementarmente, Tuytens e Devos (2012) e Abdullah e Abraham (2012) afirmam que a percepção de que o sistema de GD é justo auxilia na sua aceitação, podendo resultar em uma aceitação mais elevada do sistema por parte dos funcionários, o que é apontado por resultar em maior comprometimento organizacional (Atitudes no trabalho/Transferência) e satisfação com o trabalho (Atitudes no trabalho/Transferência) (Abdullah & Abraham, 2012).

Outra relação identificada indica que sistemas justos (Justiça/Reação) estão relacionados ao fornecimento de feedback útil (Habilidade/Aprendizagem), como descrito por Abdullah e Abraham (2012). Segundo Culberstson et al. (2013), a satisfação com GD é considerada um dos critérios de Reação que mais indica as consequências do feedback referente a um processo de avaliação, tendo em vista a relação estabelecida para aumentar motivação (Motivação/Transferência), comprometimento (Atitudes no trabalho/Transferência) e desempenho (Desempenho no trabalho/Transferência).

Dessa forma, a literatura apresenta resultados de que a dimensão Reação está diretamente relacionada à de Aprendizagem, estabelecendo relação também com a dimensão de Transferência. Porém, nesse caso, indiretamente por intermédio da dimensão de Aprendizagem. Relações diretas entre a dimensão Reação e as demais, cujo fenômeno é de nível organizacional, não foram diretamente identificadas na literatura revisada.

Alguns estudos que consideram a dimensão Reação como central ou exclusiva para avaliar a efetividade de GD (*e.g.* Cislighi & Zugno, 2019; Keeping & Levy, 2000; Modipane et al., 2019; Wright & Cheung, 2007) indicam que avaliar a efetividade de GD utilizando somente as reações que os participantes possuem acerca do sistema é possível, porém essa proposta não deve ser vista como sólida ou suficiente, pois, a partir das relações estabelecidas com outras dimensões, é possível perceber que, mesmo diante de sistemas percebidos como ruins ou errados, ainda poderão ser notados efeitos do sistema de GD em outros níveis ou outras dimensões.

É o que também afirmam Schleicher et al. (2019) sobre a adoção da dimensão Reação não ser suficiente para determinar que um sistema de GD é, ou não, fracassado, ainda que ele seja percebido como injusto, impreciso, insatisfatório, inútil e traga sentimentos e efeitos ruins. Murphy (2020) aponta que, entre as organizações, apesar da maioria de seus funcionários não

aprovarem o sistema de GD, isso não é fator suficiente para abandoná-la. O autor ainda complementa que, mesmo com a preocupação excessiva em evitar erros e tornar o sistema justo e útil, não é possível garantir os resultados desejados, pelo contrário: ele afirma que os sistemas de GD sempre serão cheios de erros, injustos e, em certa medida, inúteis. Logo, considerar apenas a dimensão Reação representa entregar uma avaliação parcial da efetividade de GD.

No caso da dimensão Aprendizagem, percebe-se que os autores Schleicher et al. (2019) buscaram explicar o mecanismo pelo qual as reações se transformam em outros resultados, uma vez que, para os autores, somente elas não deveriam manter relação suficiente para precisar o que pode ter sido adquirido como resultado da participação no sistema de GD, conforme apontado anteriormente. Dessa forma, é compreensível a lógica em que, mesmo que a efetividade de GD possa ser compreendida pelas reações, como há uma grande distância da dimensão Reação para compreender de que forma os resultados de GD são transferidos para o trabalho, o caminho causal das relações exige que seja tratada também a dimensão Aprendizagem para explicitar a efetividade de GD.

Em geral, a revisão identificou que os critérios dessa dimensão tratam sobre resultados decorrentes da participação e da interação do funcionário no sistema de GD: obtenção de clareza com relação às tarefas e funções desempenhadas, compreensão dos benefícios do sistema de GD, desenvolvimento de técnicas de fornecimento e recebimento de feedbacks mais efetivos, mudanças na intenção ou propensão de se comportar no trabalho e outros aspectos que são tidos como resultados individuais que se relacionam à GD. Logo, essa dimensão é composta por três critérios identificados: Atitude, Habilidade e Conhecimento.

Haines e St-Onge (2012) identificam que o desenvolvimento de habilidades pelos gestores relativas ao papel desempenhado por eles no processo de GD pode garantir que os resultados do sistema de GD sejam mais facilmente alcançados, apontando que um sistema de GD é percebido com uma efetividade maior quando há uma formação mais extensa e aprofundada dos avaliadores que favoreça o desenvolvimento de habilidades (Habilidades/Aprendizagem). O questionário aplicado pelos autores mensurou em que medida os gestores recebem treinamento sobre conteúdos relacionados ao planejamento de desempenho, entrega de feedback e outros tópicos semelhantes em que elevados índices de resposta refletem um treinamento mais extensivo sobre diferentes conteúdos do sistema de GD. Segundo o estudo de Lawler (2003), a partir do sistema de GD é possível desenvolver habilidades, as quais levam ao alcance de um desempenho individual superior (Desempenho no trabalho/Transferência).

Segundo o modelo teórico de Ikramullah et al. (2016), a efetividade de um sistema de GD pode ser avaliada a partir de mudanças nas ações e atitudes dos funcionários, o que pode levar a melhores resultados organizacionais (Resultados Operacionais e Resultados Financeiros). Segundo Ikramullah et al. (2016), os avaliadores são incentivados a buscar a participação dos avaliados no processo de GD para que, assim, as decisões sejam tomadas da melhor forma. Para os autores, isso não apenas cria consenso, mas também diminui as chances de conflitos (Relação interpessoal/Transferência) e, conseqüentemente, os avaliados ficam mais satisfeitos no trabalho (Atitudes no trabalho/Transferência) e comprometidos com sua organização (Atitudes no trabalho/Transferência).

Os autores Abdullah e Abraham (2012) identificaram que a aceitação (Atitude/Aprendizagem) do sistema de GD é influenciada pelo quanto os envolvidos percebem o sistema como justo (Justiça/Reação), o que ocasiona em maior compromisso organizacional, satisfação com o trabalho e melhorias na confiança no trabalho entre o superior e supervisionado, critérios considerados na dimensão seguinte, de Transferência. Os autores também relacionaram que quanto maior a clareza sobre o sistema de GD, acesso às informações, ao propósito do sistema, aos resultados de desempenho, mais propenso ele estará a desenvolver uma atitude positiva perante o sistema de GD.

O desenvolvimento de habilidades (Habilidades/Aprendizagem) de gestores de linha por meio do sistema de GD foi objeto central de estudo de Van Waeyenberg e Decramer (2018), em que os autores afirmaram que tais habilidades impactaram positivamente no comprometimento organizacional (Atitudes no trabalho/Transferência), na execução no trabalho (Desempenho no trabalho/Transferência) e na intenção de ficar na organização (Atitude/Aprendizagem). No mesmo estudo, os autores afirmam que possuir habilidades e informações claras sobre o sistema de GD e sobre o desempenho esperado pela organização irá impactar positivamente na satisfação com o sistema de GD (Satisfação/Reação).

Ao estabelecer tais relações entre os critérios de Aprendizagem, é perceptível uma relação causal que parte dos Conhecimentos, pela obtenção de informações suficientes a sua aplicação, chamada de Habilidades, e expressos a partir de intenções de agir ou se comportar, representado pelas Atitudes.

A aceitação do sistema também é apontada como um elemento central na garantia de outros resultados de GD, relacionados ao trabalho, como motivação (Motivação/Transferência), produtividade (Produtividade/Resultados Operacionais) e redução do absenteísmo (Rotatividade/Resultados operacionais) (Roberts, 1992). Logo, aceitar o sistema de GD foi

relacionado pelos autores por influenciar em resultados percebidos no desempenho do seu trabalho e que podem refletir nos resultados operacionais da organização.

Em resumo, a dimensão de Aprendizagem se mostra importante por ser a primeira a realmente identificar os resultados mais diretos e imediatos do sistema de GD, sendo antecedida diretamente pela dimensão Reação. A Aprendizagem está relacionada de forma direta à Transferência e indiretamente aos resultados de nível organizacional, como os Resultados Operacionais e os Resultados Financeiros.

2.4.2 Aprendizagem → Transferência

Acerca da dimensão de Transferência, ela é relativa ao retorno dos resultados para o trabalho, uma vez que tanto a dimensão Reação como a Aprendizagem aferem resultados relacionados ao contexto e à operacionalização do sistema de GD, enquanto os critérios de Transferência buscam demonstrar como os resultados multidimensionais do sistema de GD são expressos no trabalho. Dessa forma, foram identificados cinco critérios relativos a essa dimensão: Desempenho no trabalho, Motivação, Atitudes no trabalho, Relação interpessoal e Bem-estar.

A maioria dos autores que adotaram essa dimensão para avaliar a efetividade do sistema de GD referiram-se ao desempenho no trabalho, o qual é definido como Comportamentos expressos no trabalho voltados para o alcance de resultados organizacionais (Campbell, 1993 *apud* Motowildo et al., 1997). Esse critério possui um papel importante na avaliação de efetividade de GD, pois ele está diretamente relacionado ao propósito principal do sistema, de promover desempenho individual alinhado aos objetivos organizacionais de forma que o desempenho organizacional seja promovido, conforme afirma Aguinis (2013).

O desempenho no trabalho (Transferência) é relacionado à aceitação do sistema de GD (Atitude/Aprendizagem), desenvolvimento de habilidades (Habilidades/Aprendizagem), motivação dos funcionários (Motivação/Transferência), aumento da produtividade (Produtividade/Resultados Operacionais) e redução do absenteísmo (Rotatividade/Resultados Operacionais) (Roberts, 1995). Para os autores Biulchi e Pauli (2012), em seu estudo sobre o sistema de GD em uma organização pública de seguridade social no Brasil, a motivação (Motivação/Transferência) é apontada por impactar positivamente na produtividade (Produtividade/Resultados Operacionais) e, assim, entregar serviços de qualidade ao cidadão que o deixem satisfeito (Satisfação do público-alvo/Impacto Societal). Os resultados de GD estão altamente relacionados à motivação do funcionário, sendo que a motivação pode auxiliar

a alcançar objetivos organizacionais e reduzir taxas de rotatividade (Rotatividade/Resultados Operacionais) (Mok & Leong, 2021).

Outra relação estabelecida consiste no impacto de feedbacks efetivos (Habilidade/Aprendizagem) na promoção da motivação (Motivação/Transferência): Lawler (2003) destaca que reconhecer sistematicamente é efetivo em promover motivação no trabalho. Dewettinck (2008) também destaca o papel central do feedback intermediário entre supervisores e empregados para promover a motivação dos funcionários. Soltani et al. (2003) também ressaltaram a importância de feedbacks direcionados para motivar os funcionários, evitando interpretações voltadas para autoimagens de sucesso ou de fracasso. Iqbal et al. (2015) afirmam que a escassez de estudos sobre motivação do funcionário e feedback é a razão para sistemas de GD ineficazes. Segundo Awan et al. (2020), feedback adequado é capaz de promover engajamento no trabalho. Os autores dos estudos revisados também apontaram enquanto resultado a satisfação com o trabalho em que são expressos sentimentos positivos e contentamento com o trabalho (*e.g.* Kakkar et al., 2020; Mok & Leong, 2021; Schleicher et al., 2019; Upadhyay et al., 2020).

Outra relação estabelecida consiste na compreensão de que o sistema de GD faz parte de um contexto social em que os membros da organização podem enxergar o sistema de GD como sendo capaz de promover crescimento profissional e bem-estar, impactando positivamente em resultados individuais, como as atitudes no trabalho (satisfação e comprometimento) (Atitudes no trabalho/Transferência), e na intenção de deixar a organização (Atitude/Aprendizagem) (Kakkar et al., 2020).

2.4.3 Transferência → Resultados Operacionais

Tendo em vista essas características dos estudos de Transferência, esta dimensão é precedida pelas dimensões de Reação e Aprendizagem, sendo a menos próxima do sistema de GD das dimensões apresentadas até aqui. Há uma relação direta com os Resultados Operacionais, enquanto com o Impacto Social foi estabelecida uma relação indireta, mas não foi feita menção a nenhum impacto nos Resultados Financeiros.

2.4.4 Resultados Operacionais → Resultados Financeiros

A quarta dimensão que compõe a avaliação de efetividade de GD consiste nos Resultados Operacionais, a primeira categoria proposta a considerar resultados cujo fenômeno acontece a nível de unidade/organização e não mais individual. Nesse sentido, essa dimensão se mostra mais distal do sistema de GD e, em razão disso, seus resultados podem parecer menos acessíveis pelos atores organizacionais.

Entretanto, mostra-se bastante esclarecedor para pesquisadores e convincente para os gerentes encontrar um impacto significativo das práticas e políticas de GD em resultados organizacionais, mesmo que estes sejam mais distantes e possuam mais interferências (Colakoglu et al., 2006). Dessa forma, os critérios identificados como componentes desta dimensão foram: Produtividade, Rotatividade, Inovação e Força de Trabalho.

Os Resultados Operacionais, enquanto dimensão da efetividade do sistema de GD, são referentes ao aumento da produtividade nas organizações, o que, segundo Nankervis et al. (2012), teria relação positiva na percepção de melhora na qualidade dos serviços prestados à população, tendo em vista que era uma organização pública (Satisfação do público-alvo/Impacto Societal). Outra relação estabelecida consiste na avaliação da taxa de rotatividade (Rotatividade/Resultados Operacionais) enquanto critério de avaliação de efetividade de GD, o qual fez relação com sentimentos de felicidade ou sofrimento no trabalho (Bem-estar/Transferência) diante da possibilidade de recompensa ou sanções a partir do sistema de GD, como possíveis demissões (*e.g.* Espejo et al., 2021).

No caso dos Resultados Operacionais, as relações estabelecidas com as demais dimensões de nível unidade/organizacional não foram tão exploradas pela literatura, o que retrata o que fora apontado por Ikramullah et al. (2016) sobre quando um campo prioriza determinadas dimensões e desconsidera outras, o que acaba subvertendo conteúdos que poderiam promover novas discussões.

Dessa forma, a dimensão de Resultados Operacionais está posicionada após a dimensão Transferência, de forma direta, e faz relação indireta com o Impacto Social. A relação estabelecida com os Resultados Financeiros não foi identificada de forma explícita nos estudos como visto nos demais casos, pois, além de ter poucos estudos que mencionam os Resultados Financeiros, aqueles que assim fizeram apontam ambos os resultados (operacionais e financeiros) de forma conjunta como sendo “resultados organizacionais” (*e.g.* Nankervis et al., 2012; Stanton; Nankervis, 2011). Entretanto, com base no modelo integrativo de Schleicher et al. (2019) e na literatura de GP (*e.g.* Dyer & Reeves, 1985), entende-se que os resultados operacionais estabelecem relação direta com os resultados financeiros.

2.4.5 Resultados Financeiros → Impacto Societal

A penúltima e sexta dimensão identificada consiste nos Resultados Financeiros, a qual é composta pelos três critérios identificados: Lucratividade, Participação no mercado e Retorno financeiro. Essa dimensão ainda é pouco abordada pela literatura revisada de GD, conforme é possível inferir pela frequência de estudos que avaliaram os resultados do sistema de GD, os quais também não realizaram um estudo aprofundado dessa dimensão. A sua presença reduzida na avaliação de resultados do sistema de GD também foi identificada na revisão realizada por Schleicher et al. (2019). Logo, as relações entre essa dimensão e as demais se tornaram de difícil identificação, restando apenas o reconhecimento das relações estabelecidas pelos estudos de GP (*e.g.* Dyer & Reeves, 1995; Huselid, 1995). Portanto, a dimensão de Resultados Financeiros é precedida indiretamente pelas outras dimensões descritas até aqui, sendo apenas de forma direta pelos Resultados Operacionais.

Por último, tem-se a sexta dimensão identificada como Impacto Societal, composta somente pelo critério de Satisfação do público-alvo, cujo fenômeno é de nível ambiental, a única dimensão a considerar aspectos externos à organização para avaliação de efetividade de GD, sendo impactada indiretamente por todas as demais dimensões referentes aos aspectos internos da organização.

O estudo de Biulchi e Pauli (2012) identificou a redução do tempo de espera no atendimento do serviço público (Produtividade/Resultados Operacionais), o que concedeu segurança à população quanto à realização dos serviços de qualidade na região onde o estudo foi realizado (Satisfação do público-alvo/Impacto Societal). Já Pinto e Behr (2015) relatam que a avaliação de desempenho é vista como condição de melhoria para o serviço público.

Por fim, chama a atenção o fato de essa dimensão ser desconsiderada por grande parte dos estudos revisados, tendo figurado em apenas quatro estudos, sendo um apontado apenas teoricamente. É alarmante que tal dimensão não apareça, visto que uma avaliação de efetividade vai além das mudanças a nível individual e interno da organização, alcançando a promoção de melhorias no nível social em que a organização está inserida, conforme visto até aqui. Tal desconsideração pode estar atrelada à distância dessa dimensão aos sistemas de GD, o que torna sua avaliação estatisticamente arriscada sem a aplicação de técnicas rigorosas que consigam estabelecer mínimas relações causais.

Compreende-se o sistema de GD como uma ferramenta que, por mais que vise em última instância o desempenho organizacional, atua predominantemente na esfera comportamental, no máximo, da produtividade dos trabalhadores. Após isso, outras variáveis passam a ser inseridas, tanto individuais como contextuais, que irão afetar a relação entre as dimensões identificadas como fenômenos individuais e aqueles de nível unidade/organizacional (operacionais, financeiros e sociais), aumentando a dificuldade de mensuração direta de causas e efeitos.

Logo, conforme o modelo teórico-integrativo apresentado e as relações entre as variáveis anteriormente discutidas, mostra-se imperativo compreender ao menos se e de que forma as dimensões de nível individual – Reação, Aprendizagem e Transferência – estabelecem relação entre si a partir da percepção de colaboradores que já interagiram com o sistema de GD. Essa é uma das proposições recentemente formulada nos estudos de Schleicher et al. (2019) no campo de GD, mas que dominou as discussões científicas em outras áreas de GP, como de treinamento e desenvolvimento, sobretudo no final do século XX (*e.g.* Borges-Andrade & Abbad, 1996; Abbad et al., 2001).

À época, estudos de avaliação de treinamento consideravam a relação entre os níveis de reação, aprendizagem, comportamento no cargo e resultados organizacionais a mais aceita e frequentemente testada (Kaufman & Keller, 1994). Entretanto, resultados inconclusivos ou aparentemente inconsistentes impediam confirmar que tais níveis mantinham entre si forte relacionamento positivo, mobilizando uma quantidade elevada de estudos no campo naquele período (Abbad et al, 2001).

Acontece que no campo de GD proposições semelhantes a estas, com tentativas de explicar a relação entre estas variáveis nem mesmo chegaram a ser testadas no campo desta forma, conforme observado na revisão de literatura. A interação de outras variáveis no modelo teórico-integrativo, como contextuais, ressalta a complexidade da mensuração direta das relações de causa e efeito, evidenciando a necessidade de métodos mais robustos, como estudos experimentais ou o aprimoramento de estudos correlacionais, para explorar essas interações de maneira mais precisa. Sendo assim, é importante que se possa compreender de que forma a efetividade de GD se dá à princípio da avaliação das variáveis de nível individual. Assim, a identificação de que a Reação influencia diretamente Aprendizagem, e que Aprendizagem, por sua vez, precede Transferência permite compreender como mudanças no comportamento individual podem ser identificadas e analisadas para validar o impacto do sistema de GD. Portanto, propõe-se duas hipóteses:

H1: Reação influencia positivamente a Aprendizagem.

H2: Aprendizagem influencia positivamente a Transferência.

Sendo assim, torna-se necessário abordar a questão das medidas utilizadas para avaliar tais variáveis. A predominância de estudos com medidas perceptuais aplicadas no nível individual figura como um dos principais achados da revisão. Isso corrobora estudos anteriores que indicam a preferência por medidas de nível individual para avaliar a efetividade de Gestão de Desempenho (GD), mesmo quando as dimensões ou critérios avaliados são de nível organizacional, o que justificaria o uso de medidas também nesse nível (*e.g.*, Den Hartog et al., 2004; Guest, 1998; Keeping & Levy, 2000; Levy & Williams, 2004; Schleicher et al., 2019). No entanto, medidas de natureza objetiva são mais frequentes na dimensão “Resultados Operacionais”, aplicadas apenas a critérios como produtividade e rotatividade, embora seja possível utilizar indicadores objetivos em outros critérios dessa dimensão e das demais.

A discussão sobre a utilização de dados objetivos ou subjetivos na avaliação de resultados ainda é um tema aberto na literatura (cf. Gerhart; Fang, 2015; Wall et al., 2004). Boselie et al. (2005), por exemplo, afirma que dados objetivos de desempenho, apesar de serem mais difíceis de serem acessados, devem ser preferidos em detrimento do julgamento subjetivo dos respondentes em determinadas situações.

Wall et al. (2004) destacam que os estudos realizados no início do século adotaram dados objetivos ou subjetivos como equivalentes, mas eles não devem ser interpretados dessa forma, pois isso significa desconsiderar as características específicas ao se adotar medidas subjetivas ou objetivas. Os autores destacam que as medidas subjetivas normalmente são construídas a partir uma série de itens que são agregados para compor uma escala, baseada em um conceito de desempenho ou de resultados utilizado e multifacetado, enquanto dados objetivos utilizam dois ou três índices mais específicos e precisos e apresentam resultados para cada um separadamente, como produtividade, lucro ou retorno de investimentos. Entretanto, eles concluem que sempre que for possível a investigação dos resultados ou da efetividade de uma política ou prática organizacional, deve-se considerar medidas subjetivas e objetivas. Dessa forma, a avaliação da efetividade de políticas ou práticas organizacionais deve considerar ambos os tipos de medidas sempre que possível, buscando evitar uma predominância exclusiva de medidas subjetivas.

A noção de que muitas pesquisas têm que recorrer à percepção de respondentes para estimar resultados sobre GP é apontada por Boselie et al. (2005) em seu estudo, o qual demonstra este como sendo um padrão na avaliação de resultado no campo de políticas e práticas de GEP. Esse aspecto se relaciona também com a questão de conveniência de acesso

aos dados: a predominância entre os estudos revisados por medidas cuja natureza é subjetiva (*soft*), como a percepção, a nível individual, é apontada por Guest (1997) como a solução restante para quando não se tem acesso aos dados tidos como objetivos ou “duros” (*hard*), sobretudo aqueles relacionados aos fenômenos de nível de unidade/organizacional, e não estão disponíveis nas organizações para acesso.

Essa perspectiva dialoga com o conceito de “psicologização” na literatura de GP, descrito como o foco em explicações micro-organizacionais e psicossociais para compreender a efetividade de sistemas de GP (Budd, 2020; Godard, 2014, 2020; Kaufman, 2020). Kaufman (2020) destaca que a psicologização decorre da forte influência de teorias de comportamento organizacional e psicologia industrial-organizacional, as quais orientam padrões metodológicos mais rigorosos, como é exigido por periódicos de alto impacto (Farndale et al., 2020).

É nesse ponto que se deve equilibrar a crítica à psicologização com uma justificativa metodológica que reconheça sua relevância. No caso desse estudo, a adoção de medidas subjetivas foi uma solução metodológica necessária, visto o difícil acesso a dados objetivos. Este trabalho buscou investigar o impacto psicológico do sistema de GD que pode futuramente aprimorar o estabelecimento de conexões entre esses impactos e resultados concretos.

Tendo em vista a dispersão significativa no campo, especialmente nas dimensões de Reação, Aprendizagem e Transferência, percebeu-se o uso de instrumentos perceptuais baseados em medidas genéricas ou conceitualmente inconsistentes, dificultando uma análise robusta dessas dimensões (Schleicher et al., 2019). Além disso, a qualidade psicométrica das escalas existentes varia consideravelmente, o que evidencia a necessidade de desenvolver e validar instrumentos mais sólidos para avaliar essas dimensões. Nesse sentido, havendo essa profusão de conceitos e medidas que dificultam a mensuração de tais dimensões, a construção e validação de medidas que busquem corrigir problemas conceituais e sanar eventuais confusões operacionais tornam-se essenciais, conforme Murphy e Cleveland afirmaram em 1995, e Schleicher et al. (2019) também apontaram recentemente.

Com base nisso, torna-se possível testar adequadamente as hipóteses previamente propostas, que operacionalizam parte das relações delineadas no modelo teórico-integrativo de avaliação da efetividade da gestão de desempenho, agora adaptado à realidade da Administração Pública brasileira.

Na próxima seção, serão apresentados os métodos, técnicas e procedimentos de pesquisa utilizados para operacionalizar o modelo e testar as hipóteses formuladas neste estudo.

3 MÉTODOS, TÉCNICAS E PROCEDIMENTOS DE PESQUISA

A presente seção dispõe sobre os métodos, procedimentos e técnicas de pesquisa utilizados para a consecução dos objetivos deste estudo, descrevendo as características do estudo, as organizações participantes e os métodos e técnicas de coleta e análise de dados que serão empregados.

3.1 Características da pesquisa

Esta pesquisa teve como objetivo geral testar empiricamente um modelo de avaliação de efetividade do sistema de Gestão de Desempenho adaptado ao contexto da Administração Pública Federal brasileira. Dessa forma, classificou-se este estudo como do tipo teórico-empírico e de natureza descritivo. Dado que as informações serão obtidas em um certo momento do tempo, o recorte temporal deste estudo configura-se como transversal (Cooper & Schindler, 2014), sem intenção de que o resultado possa aferir uma evolução histórica ou mudanças ocorridas ao longo do tempo.

Este estudo empregou abordagem mista com método de pesquisa de levantamento de opinião por meio de técnicas de aplicação de questionário e de entrevista semiestruturada com servidores públicos de organizações da Administração Pública federal.

Para consecução do objetivo específico de demonstrar empiricamente as relações entre as dimensões utilizadas para avaliar a efetividade do sistema de Gestão de Desempenho no contexto da Administração Pública brasileira, a operacionalização do processo de amostragem, a construção e validação do questionário para reunião de dados, os procedimentos de coleta e, por último, as técnicas de análise dos dados foram dispostos a seguir em dois estudos: Estudo 1, referente à construção teórica e evidências de validade das escalas das variáveis individuais: Reação, Aprendizagem e Transferência, e Estudo 2, referente ao teste das relações hipotéticas entre essas variáveis individual. As operações e os procedimentos metodológicos realizados em cada estudo estão contidos na Tabela 15.

Tabela 15

Resumo das características e procedimentos metodológicos da pesquisa

Estudo	Fase	Subfase	Objetivos metodológicos	Instrumentos	Amostragem	Amostra	Análise
1	A	1A	Construir teoricamente as escalas das variáveis de perspectiva individual	Pesquisa documental	-	58 artigos	Análise de conteúdo
				Roteiro misto (semiestruturado e estruturado) de entrevista	Rígida, com tamanho e perfil pré-determinados	14 gestores ou servidores públicos de unidade de GP	Análise lexicográfica Análise de conteúdo
		2A	Validar escalas das variáveis da perspectiva individual com juízes	3 Questionários – Escala de Reação, Escala de Aprendizagem e Escala de Transferência	Não probabilística por acessibilidade	9 doutores para análise de juízes	Análise da taxa de concordância entre juízes Análise semântica dos itens do questionário
	B	1B	Apresentar evidências de validade exploratória das escalas das variáveis de perspectiva individual	Questionário de Avaliação da Efetividade do Sistema de GD	Não probabilística por acessibilidade com número determinado em razão da técnica estatística empregada	238 servidores públicos	Análise Fatorial Exploratória (AFE)
		2B	Apresentar evidências de validade confirmatória das escalas das variáveis de perspectiva individual	Questionário de Avaliação da Efetividade do Sistema de GD - após análise exploratória	Não probabilística por acessibilidade com número determinado em razão da técnica estatística empregada	145 servidores federais de duas organizações públicas	Análise Fatorial Confirmatória (AFC)
2	1	1A	Testar as relações entre as variáveis da perspectiva individual				Regressão linear

3.2 Percurso Metodológico do Estudo 1

O Estudo 1 foi desenvolvido a partir de duas fases: **Fase A**, que consiste na construção teórica das medidas das dimensões Reação, Aprendizagem e Transferência e na validação de juízes; e a **Fase B**, referente à validação empírica, com análise fatorial exploratória e confirmatória das escalas construídas anteriormente.

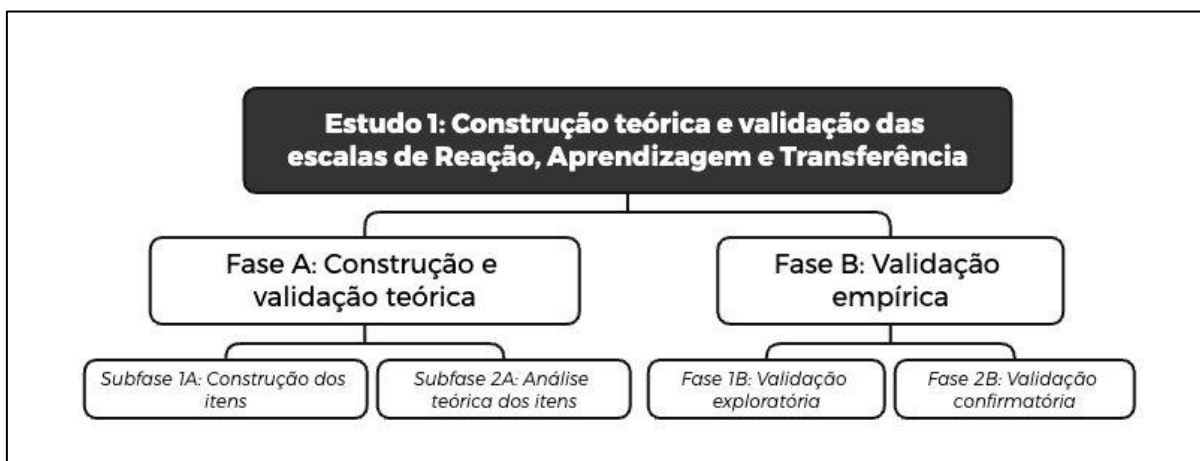


Figura 6. Fluxo de atividades metodológicas do Estudo 1

Fonte: Elaborado pela autora.

Assim, as subseções adiante dispõem acerca da consecução dessas respectivas fases.

Fase A: Construção teórica e validação de juízes das Escalas de Percepção de Reação, de Aprendizagem e de Transferência

Subfase 1A: Construção das escalas

Para construção de um instrumento de uma medida, Pasquali (2010) afirma que existem três polos que guiam os procedimentos desta pesquisa: teórico, empírico e analítico. O polo teórico é referente às teorias do campo, que orientam o construto para o qual será desenvolvido o instrumento da medida, iniciando com o estabelecimento do sistema ou variável a ser medida, seguido do desenvolvimento da teoria psicológica sobre esse traço, da operacionalização do traço através da elaboração dos comportamentos que o representam (elaboração dos itens) e da análise teórica dos itens (análise de juízes e

análise semântica). O segundo polo, empírico, refere-se às técnicas e às etapas da aplicação do instrumento que permite avaliar a qualidade psicométrica da escala. Por último, o polo analítico prevê os procedimentos de análises estatísticas efetuados para demonstração da sua validade.

Nessa subfase é descrita a operacionalização do polo teórico no que tange à elaboração e adaptação de itens para o instrumento das medidas de Reação, Aprendizagem e Transferência.

Elaboração dos itens – Subfase 1A

Inicialmente, prosseguiu-se com a construção do instrumento de maneira teórica reunindo as medidas e seus respectivos instrumentos identificados na literatura para cada dimensão (Reação, Aprendizagem e Transferência) com base nos dados oriundos da revisão de literatura empreendida anteriormente – e apresentada no capítulo anterior – e de uma busca feita a partir de bola de neve desses artigos revisados. Ao todo foram analisados 27 instrumentos que se mostraram pertinentes para a adaptação dos itens.

Em seguida, foi iniciado o processo de adaptação do instrumento das medidas reunidas à realidade brasileira. Para isso, foi necessário realizar a tradução dos itens da língua inglesa para língua portuguesa. Essa tradução foi feita inicialmente de forma literal por software de tradução, seguido pela eventual substituição de determinadas palavras por parte da autora, fluente em inglês, que se fez necessário para tornar o item compreensível. A tradução realizada seguiu princípios semânticos, idiomáticos, linguísticos, contextuais e conceituais de modo a encontrar uma única versão, conforme sugeridos por Borsa, Damásio e Bandeira (2012). As respectivas escalas provenientes da adaptação podem ser conferidas na Tabela 16.

Tabela 16

Medidas e instrumentos revisados para construção do instrumento

Critério	Instrumentos de referência	<i>n</i> de instrumentos	<i>n</i> inicial de itens traduzidos
Reação		10	106

Justiça	<i>Fairness Items</i> (Kim; Rubianty, 2011); Heneman; Milanowski (2004); Harrington (2015); <i>Justice Measure Items</i> (Colquitt, 2001)		37
Precisão	Heneman; Milanowski (2004); <i>PMS accuracy scale items</i> (Sharma et al., 2016)		15
Satisfação	Decramer et al. (2012); Heneman; Milanowski (2004); Kuvaas (2006); <i>Performance Appraisal Satisfaction</i> (Katevich, 2013); Youngcourt et al. (2007)		42
Utilidade	Greller (1987); Heneman; Milanowski (2004); Tuytens & Devos (2012)		7
Afetividade	Greller (1978); Heneman; Milanowski (2004)		5
Aprendizagem			39
Habilidade	Dewettinck & van Dijk (2013); Gilbert et al. (2011); Kuvaas (2011)	5	19
Atitude	Guay, Vallerand; Blanchard (2000)		11
Conhecimento	Gilbert et al. (2011); Rizzo, House e Lirtzman (1970)		9
Aprendizagem			66
Desempenho no trabalho	Brockner et al. (1992); Dewettinck & van Dijk (2013); May et al., 2002		8
Atitudes no trabalho	Michigan organizational assessment questionnaire (Cammann et al., 1983); Tremble and Alderks (1992); <i>Utrecht Work Engagement Scale</i> (Martins, 2013); Youngcourt et al. (2007)	11	31
Motivação no trabalho	Cameron and Pierce (1994); Gagné et al. (2010)		19
Relação interpessoal	Rizzo et al. (1970) - <i>Perceived supervisor support (PSS)</i>		5
Bem-estar	Zheng et al. (2015)		3

Com a versão preliminar dos itens traduzidos, esta foi submetida à análise teórica com a população para qual o instrumento foi construído com intuito de analisar a compreensão, a magnitude e a pertinência dos itens (Cassepp-Borges et al., 2010; Pasquali, 2011). Essa etapa também se mostrou útil para construção do instrumento por permitir identificar eventuais ajustes nos itens diante da sua adaptação à realidade da Administração Pública brasileira.

Para isso, foi feita uma consulta a gestores e servidores de unidades de Gestão de Pessoas de órgãos da Administração Pública federal (APF) com contato prévio com o sistema de Gestão de Desempenho da organização. Diante do universo de 204 organizações da APF, buscou-se selecionar os participantes oriundos de órgãos de natureza jurídica distintas, priorizando aspectos como a diversidade e a variabilidade, os

quais são importantes para a abordagem qualitativa desta etapa da pesquisa (Flick, 2009). Seguindo esses princípios, foram contatados mais de 25 organizações e mais de 30 atores a partir da disponibilidade dos contatos em meio público ou privado ou pela indicação de contatos dos próprios participantes da pesquisa.

Os participantes foram formalmente convidados por meio de eletrônico ou por telefone, cuja mensagem continha as informações necessárias para participação na pesquisa, conforme consta na Carta de Apresentação (Apêndice A). Nesse sentido, os participantes confirmaram sua participação gradualmente e a consulta a novos participantes foi interrompida quando os dados obtidos não permitiram mais a geração de novas informações ou *insights* para mais explorações do conteúdo. Esse processo de amostragem é chamado de saturação teórica, que ocorre quando a incorporação de novos casos para subsidiar a teorização almejada não são mais depreendidos do campo de observação (Fontanella et al., 2011). Ao todo, 14 servidores de 11 órgãos aceitaram contribuir para a pesquisa, conforme disposto na Tabela 17.

Tabela 17

Participantes das entrevistas

<i>n</i> entrevistas	Organização	Tipo	Cargo	<i>n</i> entrevistados
1	Organização 1	Ministério	Diretor de Gestão de Pessoas	1
2	Organização 2	Fundação	Chefe da Divisão de Desempenho	1
3	Organização 3	Presidência	Coordenador de Gestão de Pessoas	1
4	Organização 4	Ministério	Coordenador-Geral de Planejamento da Força de Trabalho	1
5	Organização 5	Autarquia	Coordenador-Geral de Gestão de Pessoas	1
6	Organização 6	Autarquia	Coordenador-Geral de Gestão de Pessoas	1
7	Organização 7	Agência reguladora	Superintendente de Gestão de Pessoas Coordenador	2
8	Organização 8	Empresa pública	Coordenador de Gestão de Pessoas	1
9	Organização 9	Agência reguladora	Superintendente de Gestão Estratégica de Pessoas Gerente de Gestão Estratégica de Pessoas	2

10	Organização 10	Ministério	Coordenador-Geral de Gestão de Desempenho	1
11	Organização 11	Autarquia	Coordenador-Geral de Gestão Estratégica de Pessoas Chefe de Divisão de Capacitação	2

Nota: Todos os cargos foram colocados sem menção ao gênero do ocupante, de forma a garantir o anonimato.

Dessa forma, foram realizadas 11 entrevistas em meio eletrônico. As entrevistas foram iniciadas com a leitura do Termo de Consentimento Livre e Esclarecido com intuito de reafirmar as motivações da pesquisa anteriormente apresentadas e garantir a confidencialidade dos dados, a utilização das informações exclusivamente para fins da pesquisa científica e o consentimento dos participantes com relação ao registro eletrônico da imagem e do áudio, sendo todas as entrevistas gravadas para posterior análise dos dados.

As entrevistas foram conduzidas a partir de instrumento de pesquisa composto por perguntas semiestruturadas, escala a ser analisada e perguntas sociodemográficas (Instrumento 1 - Apêndice B). Dessa forma, o roteiro de pesquisa semiestruturado foi composto por três seções: a Seção 1 contém três perguntas sobre as características gerais do sistema de GD na organização do participante, visando identificar antecedentes e consequentes tidos como importantes em um sistema de GD pelo participante. A Seção 2 inicialmente introduz o participante no contexto dos critérios e dimensões adotados pela literatura para avaliar a efetividade de GD e propõe quatro perguntas norteadoras sobre como esses aspectos aparecem na organização do participante em relação ao sistema de GD da organização; dessa forma torna-se possível capturar as percepções dos servidores da organização e do entrevistado sobre o sistema de GD da organização. Por último, a Seção 3 é composta pela escala a ser analisada pelos entrevistados. Nesse caso, cada entrevistado analisou item a item, conforme orientação de Pasquali (2011), de um tipo de escala: ou de Reação ou de Aprendizagem ou de Transferência, a fim de priorizar a qualidade das contribuições em cada um dos instrumentos. Essa decisão foi tomada após a primeira entrevista, a qual também serviu de piloto para calibragem do instrumento de pesquisa, constatando-se a necessidade de otimizar o tempo das entrevistas, evitando o desgaste e a desmotivação do entrevistado. As seções e seus respectivos objetivos estão descritos na Tabela 18.

Tabela 18

Seções do Roteiro de Entrevista

Seção	Instrumento	Perguntas
Seção 1: Características gerais do sistema de GD, visando identificar antecedentes e consequentes tidos como importantes em um sistema de GD pelo participante.	Semiestruturado	1. Em geral, quais fatores ou aspectos você considera importantes em um sistema de GD/AD? 2. O que um sistema de GD/AD tem que ter ou ser para ser considerado um sistema efetivo? 3. Na sua opinião, quais resultados ou consequências decorrem de um sistema de GD/AD de sucesso?
Seção 2: Critérios e dimensões adotados pela literatura para avaliar a efetividade de GD	Semiestruturado	1. O que você teria a dizer sobre estes aspectos no sistema de GD/AD da organização? 2. Como as pessoas da organização esperam que o sistema de GD/AD seja? E o que elas dizem sobre o sistema da organização? 3. Quais são suas percepções sobre o sistema de GD/AD da organização? E dos demais servidores? 4. As pessoas normalmente visualizam quais aprendizados, resultados ou consequências decorrentes do sistema de GD/AD?
Seção 3: Classificação dos itens traduzidos para realidade organizacional	Estruturado	Escalas 1. Análise da relevância dos itens propostos para a efetividade do sistema de GD

Com a gravação e transcrição das entrevistas realizadas, foi feita a análise dos dados. Inicialmente, foi utilizado o software *Iramuteq* de forma auxiliar à análise de conteúdo. A partir do software, foi conduzida uma análise gráfica de similitude fundamentada na teoria dos grafos, permitindo identificar as relações de ocorrência entre palavras. Os resultados oferecem *insights* sobre as conexões entre os termos, facilitando a compreensão da estrutura examinada e contribuindo para a identificação das variáveis relevantes na análise de dados (Ratinaud & Marchan, 2012). Na análise do dendrograma, a compilação de termos utiliza o teste qui-quadrado (χ^2), das palavras que demonstraram um valor superior a 3,84 e $p < 0,0001$.

Os dados analisados no *Iramuteq* complementaram a análise de conteúdo pré e pós-categórica, fundamentada nos princípios de exclusão mútua, homogeneidade, pertinência, objetividade, fidelidade e produtividade, conforme os pressupostos de Bardin

(2016). A análise pré-categórica baseou-se nas dimensões e critérios previamente definidos na etapa anterior, alinhados à literatura revisada. Assim, os conteúdos das entrevistas foram organizados de acordo com as dimensões de Reação, Aprendizagem e Transferência e seus respectivos critérios. Por outro lado, a análise pós-categórica foi conduzida com o objetivo de identificar categorias não contempladas previamente pela revisão da literatura, ampliando a compreensão dos dados. Essa análise conjunta se mostrou importante por verificar se as categorias identificadas anteriormente durante a revisão de literatura e que compuseram os instrumentos faziam sentido no contexto específico deste estudo, além de permitir analisar outros conteúdos não identificados anteriormente. Esse passo foi essencial por não se acreditar ser suficiente apenas a tradução de escalas internacionais, próprias de outros contextos.

Além disso, também foi objeto de análise a compreensão dos itens traduzidos (Pasquali, 2011). Seguindo as orientações do autor, um item que não apresente dificuldade de compreensão pelos sujeitos em dois momentos distintos não necessita mais ser ulteriormente checado, porém, caso haja, deve-se realizar até seis conferências com outros sujeitos, devendo ser descartado se permanecer duvidoso. Dessa forma, nem todos os itens das escalas foram submetidos a todas os entrevistados, considerando que cada entrevistado analisou pelo menos um tipo de escala.

Ao longo das entrevistas alguns itens foram reformulados quando surgiam dúvidas na compreensão ou quando houve uma compreensão distinta daquilo que a autora da pesquisa pretendia. Quando todas as entrevistas foram finalizadas, houve a retirada de itens por duplicidade ou por não representar o atributo a que o item se refere. Além disso, foi feita a inclusão de itens por remeter a aspectos apontados pelos entrevistados e que não haviam sido abarcados pela literatura anteriormente.

Subfase 2A: Análise teórica dos itens

Dessa forma, com as escalas ajustadas após essa etapa de análise, seguiu-se para análise de conteúdo e análise semântica. Essa análise foi feita por peritos na área que ajuizaram se os itens eram referentes, ou não, ao atributo em questão, indicando em que medida os itens representavam o fator, buscando verificar o nível de concordância entre os juízes (Pasquali, 2011). Para esta etapa, foram convidados a contribuir com a pesquisa 12 (doze) juízes, porém apenas 9 (nove) aceitaram contribuir com esta etapa da pesquisa, sendo que 3 (três) juízes diferentes analisaram cada um dos instrumentos. Essa quantidade

converge com a recomendação da literatura que indica entre 3 (três) a 5 (cinco) juízes para realização da validação (Cassepp-Borges et al., 2010) ou no mínimo 2 (dois) juízes (Borsa & Seize, 2017). Todos os juízes-validadores eram doutores e especialistas em gestão de pessoas com publicações nos últimos 3 (três) anos. Nenhum dos juízes-validadores tinha tido contato anterior com esta pesquisa em outra etapa.

A participação dos juízes enquanto validadores se deu mediante convite enviado por mensagem eletrônica contendo o instrumento da escala a ser validada (Instrumento 2 – Apêndice C) em conjunto com o Termo de Consentimento Livre e Esclarecido, documentos que deveriam ser devolvidos após o seu preenchimento em resposta ao e-mail. O arquivo contendo a escala de cada dimensão de avaliação de efetividade (Reação, Aprendizagem e Transferência) foi enviado para os juízes com a aba de orientações, contendo a definição da dimensão e dos critérios, e a escala a ser validada.

A tarefa dos juízes consistia em classificar cada um dos itens a partir de dois aspectos: a) adequação do item ao fator, verificando qual faceta pertence a cada item, tendo como base a definição apresentada; e b) clareza, avaliando o quanto os itens são compreensíveis (diretos, claros e objetivos) a partir de escala do tipo *likert*, devendo sugerir adequações, caso fosse incompreensível.

Após a conclusão de todas as validações, foram calculados dois índices: a Taxa de Concordância de Juízes (TCJ) e o Coeficiente de Validade de Conteúdo (CVC). A concordância entre os juízes foi obtida comparando-se as suas avaliações. Por se tratar de uma variável categórica, utilizou-se o coeficiente *kappa* médio, obtendo-se a média entre os juízes. Como critério de exclusão, foram eliminados itens com concordância inferior a 67%, considerando o número de juízes participantes. Esse ponto de corte foi adotado com base no critério de (Landis & Koch, 1977), que classificam concordâncias próximas a 60% como moderadas.

O CVC, além de medir a concordância entre os avaliadores, utiliza uma escala do tipo *Likert* para avaliar a equivalência de conteúdo. Neste estudo, os juízes, além de atribuírem um fator (ou faceta) a cada item do questionário, avaliaram também a clareza da linguagem numa escala de 1 a 5. Foram considerados aceitáveis os itens cujo CVC fosse superior a 0,8 (Cassepp-Borges et al., 2010). Neste estudo, foram considerados itens com CVC acima de 0,70, flexibilizando ligeiramente o ponto de corte, devido à quantidade reduzida de avaliadores. Esse ajuste é viável em pesquisas, sem comprometer os resultados obtidos (Cassepp-Borges et al., 2010).

Após as análises teórica e semântica, procedeu-se às etapas de validação empírica do instrumento de pesquisa.

Fase B: Validação empírica das medidas

Os procedimentos de validade empírica desta pesquisa foram divididos em duas subfases com amostras distintas. A Subfase 2A.1 contemplou a aplicação do questionário preliminar e os procedimentos da análise fatorial exploratória (AFE). Já a Subfase 2A.2 contemplou as etapas da análise fatorial confirmatória (AFC).

Método da subfase 1B – Análise Fatorial Exploratória (AFE)

Demonstram-se, a seguir, o instrumento, os procedimentos de coleta, a amostra e a análise de dados que embasaram a análise fatorial exploratória (AFE).

Instrumento – Subfase 1B

O instrumento de pesquisa desenvolvido na etapa anterior foi disponibilizado eletronicamente por meio do software *Microsoft Forms*, permitindo acesso on-line aos participantes (Apêndice D). A estrutura do instrumento seguiu um formato para facilitar o fluxo de resposta dos participantes. Inicialmente, foi apresentada uma introdução à pesquisa, seguida pelo Termo de Consentimento Livre e Esclarecido (TCLE). O TCLE detalhou o título da pesquisa, objetivos, procedimentos, potenciais riscos e benefícios, garantias de confidencialidade, a natureza voluntária da participação, uma estimativa do tempo necessário para o preenchimento e outras informações essenciais para embasar a decisão dos participantes sobre aderir ao estudo. Ao final dessa seção, os participantes foram convidados a concordar ou discordar dos termos. Aqueles que optaram por não consentir foram automaticamente direcionados ao encerramento da pesquisa, sem que suas respostas fossem registradas.

Os participantes que concordaram com o TCLE responderam, em seguida, a uma pergunta sobre sua participação em processos de gestão ou avaliação de desempenho em organizações públicas. Participantes que responderam negativamente foram conduzidos ao encerramento da pesquisa. Já os que responderam afirmativamente foram

encaminhados para uma etapa de filtragem, onde uma segunda pergunta direcionava cada respondente a uma das três escalas (Reação, Aprendizagem ou Transferência) de acordo com o trimestre de seu aniversário: janeiro a abril para a escala de Reação, maio a agosto para a de Aprendizagem e setembro a dezembro para a de Transferência. Essa estratégia visou assegurar amostras independentes em cada escala, minimizando o risco de contaminação de dados entre as diferentes medidas.

A terceira parte do instrumento abrangeu as escalas de Reação, Aprendizagem e Transferência, com o objetivo de validar as medidas de cada constructo. A escala de Reação continha 44 itens, a de Aprendizagem, 22 itens, e a de Transferência, 33 itens. A resposta aos itens foi estruturada em uma escala de *Likert* de cinco pontos, conforme a recomendação de (Antonialli et al., 2016), que destacam sua precisão equiparável à de escalas de sete pontos. A escolha pela escala de cinco pontos também se justifica pela simplicidade e eficácia para instrumentos de grande extensão, como neste estudo. A escala de dez pontos, por sua vez, foi descartada devido à sua obsolescência, conforme apontado por Dalmoro e Vieira (2014).

Coleta de dados – Subfase 1B

A coleta de dados teve início em 31 de julho de 2023 e foi concluída em 23 de setembro de 2023. A princípio, o instrumento de pesquisa foi divulgado pelas redes sociais pessoais da pesquisadora (WhatsApp, Instagram e LinkedIn). No entanto, após três semanas e com um número de respostas abaixo do esperado, ampliou-se a estratégia de divulgação por meio de mala direta (e-mail) para as unidades de gestão de pessoas de órgãos públicos. Além disso, uma ação de panfletagem com QR Codes direcionando ao formulário foi realizada durante um congresso científico sobre Administração Pública, ocorrido em Brasília em agosto de 2023. Posteriormente, foi solicitada parceria institucional com três órgãos públicos, mas apenas um órgão aceitou participar, contribuindo substancialmente para o aumento do número de respostas.

O monitoramento da coleta de dados foi realizado semanalmente. Após quase dois meses de coleta, e com a quantidade de respostas ainda abaixo do esperado, optou-se por uma modificação na estrutura do questionário: ao invés de direcionar o respondente a uma única escala, passou-se a permitir que todos respondessem a todas as escalas (Reação, Aprendizagem e Transferência), deixando de considerar as amostras como

independentes. Essa mudança resultou em um aumento nas respostas, permitindo encerrar o período coleta de dados com 315 respostas.

Para adequar os dados às análises, foi realizada uma triagem inicial. Três respondentes foram excluídos por não terem concordado com o Termo de Consentimento Livre e Esclarecido (TCLE) e, portanto, não completaram o questionário. Além disso, foram removidos 74 respondentes que afirmaram nunca ter participado de um processo formal de avaliação ou gestão de desempenho no setor público. Todos os questionários restantes estavam completos, uma vez que o preenchimento de cada item era obrigatório, totalizando 238 respostas aptas para análise.

Amostra – Subfase 1B

O lócus da pesquisa foi a Administração Pública brasileira, e a amostragem adotada foi não probabilística por conveniência. Participaram do estudo exclusivamente servidores públicos da Administração direta ou indireta que consentiram com os termos do questionário eletrônico. Dessa forma, a amostra foi composta apenas por servidores de órgãos públicos, em alinhamento com o objetivo de avaliar a efetividade do sistema de gestão de desempenho na Administração Pública brasileira.

A quantidade ideal de respondentes para a realização da AFE é um tema de divergência na literatura, devido aos resultados inconsistentes. Estudos apontam que a qualidade de uma solução fatorial não depende exclusivamente do número de respondentes, mas também da qualidade do instrumento de medida utilizado (Damásio, 2012). Segundo MacCallum et al. (1999) soluções fatoriais estáveis podem ser alcançadas com amostras variando entre um a três respondentes por item. Esses autores demonstram que a qualidade de uma solução fatorial está mais relacionada ao grau de sobre-determinação dos fatores identificados do que ao tamanho da amostra. Sobre-determinação refere-se à representação clara de um fator por um número adequado de itens e pela qualidade das cargas fatoriais e comunalidades. Quando um fator é representado por um número suficiente de itens (geralmente quatro ou mais) e esses itens exibem altas cargas fatoriais (acima de 0,60), o tamanho da amostra torna-se menos crítico para a obtenção de uma estrutura fatorial robusta (Damásio, 2012). De acordo com MacCallum et al. (1999), o erro amostral tende a diminuir quando os fatores são fortemente sobre-determinados, tornando menos crucial um tamanho amostral elevado para análises fatoriais.

Assim, o tamanho amostral foi ajustado para o mínimo viável, considerando as dificuldades de coleta no período disponível, com base em evidências de que uma amostra menor não prejudicaria as análises, desde que houvesse ao menos dois respondentes por item.

Além do critério de quantidade de respostas por item, foi aplicado o teste Kaiser-Meyer-Olkin (KMO) para avaliar a adequação da amostra à análise fatorial. De acordo com esse critério, fatores que apresentarem valores de KMO inferiores ou próximos a 0,5 indicam a necessidade de expandir a coleta de dados. Além do aspecto quantitativo, foram considerados os critérios de diversidade e variabilidade da amostra, buscando uma representação abrangente da diversidade de órgãos públicos envolvidos (Pasquali, 2013). Para atender a esses critérios de variabilidade, a pesquisa incluiu participantes de todos os poderes – Executivo, Legislativo, Judiciário e órgãos independentes. Assim, embora a amostragem por conveniência tenha sido adotada, buscou-se garantir que as respostas refletissem a amplitude da Administração Pública brasileira.

Com base nesses critérios e excluindo as respostas indicadas anteriormente, determinou-se que a amostra de 238 participantes válidos era adequada para o estudo. A amostra final consistiu em 115 respondentes para a escala de Reação, 116 para a escala de Aprendizagem e 117 para a escala de Transferência. A Tabela 19 apresenta a caracterização da amostra, demonstrando o alinhamento com os critérios de diversidade e variabilidade estabelecidos.

Tabela 19

Características da amostra da AFE

Variáveis	<i>n</i>	%
<i>Sexo</i>		
Feminino	131	55%
Masculino	106	45%
Prefiro não responder	1	0,4%
<i>Esfera</i>		
Federal	154	64,7%
Estadual/Distrital	79	33,2%
Municipal	5	2,1%
<i>Poder Federativo</i>		
Executivo	189	79,4%
Legislativo	20	8,4%

Judiciário	19	8,0%
Independente	10	4,2%
<i>Escolaridade</i>		
Especialização	115	48,3%
Mestrado	57	23,9%
Ensino Superior	33	13,9%
Doutorado	23	9,7%
Ensino Médio	6	2,5%

Em relação ao sexo, a maioria dos participantes se identificou como do gênero feminino, representando 55% da amostra (n = 131), enquanto 45% (n = 106) se identificaram como do gênero masculino. Apenas um respondente (0,4%) optou por não declarar seu gênero. Quanto à esfera de atuação, 64,7% dos respondentes (n = 154) atuam na esfera federal, demonstrando uma predominância de servidores federais na amostra. A esfera estadual/distrital foi representada por 33,2% dos participantes (n = 79), enquanto uma minoria de 2,1% (n = 5) atua na esfera municipal.

A variável poder federativo também foi analisada, com a maioria dos respondentes (79,4%, n = 189) atuando no Poder Executivo, seguido por 8,4% (n = 20) no Legislativo, 8,0% (n = 19) no Judiciário, e 4,2% (n = 10) em órgãos independentes. Essa distribuição reflete uma amostra fortemente composta por servidores do Executivo, mas com representações relevantes dos demais poderes e órgãos.

No que tange à escolaridade, os dados indicam que a amostra é composta majoritariamente por servidores com alto nível de qualificação educacional. A maior parcela dos participantes possui especialização (48,3%, n = 115), seguida por aqueles com mestrado (23,9%, n = 57). Outros 13,9% (n = 33) possuem ensino superior completo, enquanto 9,7% (n = 23) possuem doutorado e 2,5% (n = 6) possuem ensino médio.

Essas características da amostra refletem um perfil de respondentes com elevado nível educacional e com uma significativa concentração de servidores federais e do Poder Executivo. Esse perfil é coerente com o foco da pesquisa na avaliação da política de gestão de desempenho na Administração Pública brasileira.

Análise de dados – Subfase 1B

As escalas foram validadas por meio de Análise Fatorial Exploratória (AFE) para avaliar a estrutura fatorial das três medidas: Escala de Percepção de Reação da Gestão de Desempenho (EPR-GD), Escala de Percepção de Aprendizagem da Gestão de

Desempenho (EPA-GD) e Escala de Percepção de Transferência da Gestão de Desempenho (EPT-GD).

A AFE foi conduzida no software FACTOR (versão 12.04.04), desenvolvido pela Universidade Rovira i Virgili de Tarragona, que oferece funcionalidades exclusivas, como a análise fatorial de *ranks* mínimos. Esse é o único método que permite calcular a proporção da variância explicada por cada fator, além de suportar correlações policóricas, o que é especialmente útil para dados ordinais (Lorenzo-Seva & Ferrando, 2006).

Para realizar a análise, foram adotados dois critérios principais: o uso de matriz policórica e o método de *extração Robust Diagonally Weighted Least Squares* (RDWLS) (Asparouhov & Muthén, 2010). A matriz policórica considera os dados como ordinais, sendo mais apropriada para variáveis de autorrelato (Timmerman & Lorenzo-Seva, 2011). O método RDWLS, por sua vez, é adequado para lidar com desvios de normalidade nos dados (Asparouhov & Muthén, 2010).

A decisão sobre o número de fatores a serem retidos foi baseada na técnica de Análise Paralela com permutação aleatória dos dados observados, conforme recomendado por (Timmerman & Lorenzo-Seva, 2011). Essa abordagem permite a construção de intervalos de confiança para os índices calculados, por meio da replicação dos dados originais do estudo (Lorenzo-Seva & Ferrando, 2006). A técnica utilizada, denominada “permutação otimizada dos dados”, difere da abordagem clássica de Monte Carlo proposta por Horn (1965) (Timmerman & Lorenzo-Seva, 2011).

A rotação escolhida foi a oblíqua, pressuposta pela correlação entre os fatores, com o método *Robust Promin*, uma variante do *Simplimax* que facilita as decisões de rotação (Ferrando & Lorenzo-Seva, 2018).

A estabilidade dos fatores foi avaliada por meio do índice H, que verifica a qualidade da representação de um fator comum pelos itens da escala. Valores de H variam de 0 a 1, sendo que valores acima de 0,80 indicam uma variável latente bem definida e estável em diferentes amostras, enquanto valores baixos sugerem instabilidade (Ferrando & Lorenzo-Seva, 2018).

Também foram analisados três indicadores de unidimensionalidade: (a) *Unidimensional Congruence* (UniCo), (b) *Explained Common Variance* (ECV) e (c) *Mean of Item Residual Absolute Loadings* (MIREAL). Valores de UniCo superiores a 0,95 indicam que a medida pode ser considerada unidimensional, enquanto valores de ECV acima de 0,85 e MIREAL abaixo de 0,3 reforçam essa conclusão (Ferrando & Lorenzo-Seva, 2018).

A adequação do modelo foi avaliada por meio dos índices de ajuste *Root Mean Square Error of Approximation* (RMSEA), *Comparative Fit Index* (CFI) e *Tucker-Lewis Index* (TLI). Segundo Brown (2006), o RMSEA deve ser inferior a 0,08 (idealmente menor que 0,06), com o intervalo de confiança não ultrapassando 0,10. Os valores de CFI e TLI devem ser superiores a 0,90, preferencialmente acima de 0,95 (Brown, 2015). Embora esses índices sejam comumente aplicados em Análises Fatoriais Confirmatórias, o software FACTOR permite que sejam gerados também na AFE. Por fim, de forma complementar, utilizou-se o teste de adequação Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett para verificar a adequação da amostra aos procedimentos fatoriais.

Método da subfase 2B – Análise Fatorial Confirmatória (AFC)

Demonstram-se, a seguir, o instrumento, os procedimentos de coleta, a análise de dados e a amostra que embasaram a análise fatorial confirmatória (AFC).

Instrumento – Subfase 2B

O instrumento de pesquisa utilizado para a realização da AFC consistiu em uma versão aprimorada das escalas previamente validadas na AFE, com uma quantidade reduzida de itens em cada escala. A quantidade final de itens está descrita na seção de resultados deste estudo.

O questionário foi desenvolvido em formato eletrônico e aplicado em duas plataformas distintas, uma para cada um dos órgãos parceiros da pesquisa. Para o primeiro órgão, a coleta de dados foi realizada pelo *Microsoft Forms*, enquanto para o segundo órgão, utilizou-se o *Google Forms*¹.

O questionário foi estruturado em cinco seções principais: (a) Apresentação da Pesquisa e Termo de Consentimento Livre e Esclarecido (TCLE); (b) Escala de Reação; (c) Escala de Aprendizagem; (d) Escala de Transferência; e (e) Dados Demográficos

¹O questionário não foi disponibilizado neste trabalho, pois havia menção ao órgão, o que poderia caracterizá-lo e comprometer a confidencialidade. Todos os itens do questionário desta subfase foram apresentados na seção de Resultados.

A pedido de um dos órgãos participantes, algumas adaptações foram realizadas no questionário. Adicionaram-se questões demográficas específicas, como unidade de lotação e tempo de atuação na organização. Além disso, foi solicitado que a ordem das seções fosse alterada, começando pela escala de Transferência, seguida pelas escalas de Aprendizagem e Reação.

Coleta de dados e análise de dados – Subfase 2B

A coleta de dados foi realizada especificamente em órgãos públicos, utilizando uma amostragem de conveniência e não probabilística. Participaram apenas os servidores públicos que consentiram em participar após lerem e aceitarem os termos no início do questionário eletrônico e de confirmarem que já haviam interagido com o sistema de GD do órgão.

Os sistemas de GD nestes dois órgãos apresentavam diferentes níveis de maturidade. Apesar de ambos contarem com a avaliação de desempenho para fins de progressão e promoção na carreira, cada órgão conta com iniciativas gerenciais específicas, como programas de avaliação por competências voltados para lideranças, avaliação de desempenho para estágio probatório e abordagens focadas no desenvolvimento profissional, entre outras. Cabe destacar que os participantes não foram orientados a limitar suas percepções a uma iniciativa específica do sistema de GD, mas sim a compartilhar suas impressões de forma ampla e geral.

Assim, para viabilizar essa etapa, foram realizadas tratativas no início de 2024 com dois órgãos públicos, com o objetivo de estabelecer parcerias institucionais que facilitassem e agilizassem a coleta de dados junto aos servidores do órgão. A pesquisa foi autorizada somente após a obtenção do aval institucional das áreas de recursos humanos e da alta gestão de cada órgão, com a assinatura do Termo de Aceite Institucional. Em função disso, fatores políticos e operacionais atrasaram o início da coleta, que apenas foi autorizada no segundo semestre de 2024.

Em cada um dos órgãos, um servidor atuou como “ponto focal”, auxiliando na implementação das estratégias de coleta de dados definidas em conjunto com a pesquisadora. Esse suporte foi essencial para a divulgação e coordenação do processo de coleta, assegurando a adesão dos participantes conforme planejado.

O primeiro órgão participante, pertencente ao Poder Executivo federal, autorizou a aplicação da pesquisa a todos os seus servidores a partir de julho de 2024. A coleta de

dados neste órgão ocorreu via *Microsoft Forms*, com início em 5 de agosto de 2024 e término em 23 de agosto de 2024. A população amostral desse órgão, conforme informado pelo ponto focal, consistiu em 310 servidores.

O segundo órgão participante foi do Poder Judiciário federal, que, a pedido da instituição, utilizou o *Google Forms* para a coleta, realizada entre 23 de setembro de 2024 e 7 de outubro de 2024. A divulgação nesse órgão foi realizada por e-mail, *banners* com o link da pesquisa e por redes sociais (WhatsApp) em grupos de trabalho. Além disso, o órgão do Judiciário apoiou a produção dos materiais de divulgação, com envolvimento da unidade de comunicação para a criação de materiais gráficos personalizados direcionados à pesquisa.

O monitoramento da coleta de dados foi realizado diariamente, com o objetivo de acompanhar o número de respondentes. A coleta foi encerrada com um total de 177 respostas, sendo 145 válidas – 27 provenientes do Poder Executivo federal e 118 do Poder Judiciário federal. Esses números refletem a quantidade de respostas que foram consideradas para a análise dos dados.

Os dados foram organizados em uma planilha Excel e analisados com o software JASP (versão 0.19.1), desenvolvido pela Universidade de Amsterdã, em que foi realizada a condução da Análise Fatorial Confirmatória (AFC).

A AFC e a AFE foram conduzidas utilizando o método de extração *Robust Diagonally Weighted Least Squares* (RDWLS), recomendado para lidar com dados que apresentam desvios de normalidade (Asparouhov & Muthén, 2010). O valor do *Bayesian Information Criterion* (BIC) foi calculado pelo método de extração *Maximum Likelihood* (ML), que também é apropriado para dados não normais (Brown, 2015). Embora sejam necessários mais estudos sobre os índices de ajuste em análises fatoriais, soluções fatoriais estáveis podem ser alcançadas com amostras variando entre um a três respondentes por item (MacCallum et al., 1999). Já Brown (2015) recomenda que modelos medianos (com 10 a 15 variáveis) tenham uma amostra em torno de 150 casos para resultados confiáveis. Com base nisso, considerou-se que os 145 casos coletados eram suficientes para as escalas EPR-GD, que contém 16 variáveis, a EPA-GD, com 11 variáveis, e EPT-GD, que contém 17 variáveis.

A adequação do modelo foi avaliada por meio dos índices de ajuste *Root Mean Square Error of Approximation* (RMSEA), *Standardized Root Mean Square Residual* (SRMR), *Comparative Fit Index* (CFI) e *Tucker-Lewis Index* (TLI). Para um ajuste adequado, os valores de SRMR devem ser inferiores a 0,08, enquanto os valores de

RMSEA devem ser próximos de 0,06 ou menores. Além disso, valores de CFI e TLI acima de 0,90 (preferencialmente acima de 0,95) indicam um bom ajuste do modelo (Brown, 2015). A confiabilidade dos fatores foi avaliada pelo coeficiente alfa de Cronbach, e as hipóteses de pesquisa foram novamente testadas, contribuindo para a análise de validade convergente dos constructos.

Amostra – Subfase 2B

Considerou-se a amostra (de 145 participantes, contemplando os dois órgãos) adequada para a realização da AFC, a qual está descrita na Tabela 20.

Tabela 20

Características da amostra da AFC

Variáveis	n	%
Sexo		
Feminino	71	49%
Masculino	70	48%
Prefiro não responder	4	3%
Cargo de liderança		
Não ocupa cargo de gestão	100	69%
Ocupa cargo de gestão	45	31%
Tempo no órgão		
Mais de 15 anos	66	46%
0 a 3 anos	32	22%
6 a 10 anos	21	14%
11 a 13 anos	17	12%
4 a 5 anos	9	6%
Idade		
35 a 44 anos	54	37%
45 a 55 anos	44	30%
25 a 34 anos	21	14%
55 a 64 anos	25	17%
65 anos ou mais	1	1%
18 a 24 anos	0	0%

Em relação ao sexo, a distribuição foi praticamente equilibrada: 49% dos participantes se identificaram como do gênero feminino (n = 71) e 48% como do gênero masculino (n = 70). Um pequeno percentual (3%, n = 4) optou por não declarar seu

gênero. Quanto ao cargo de liderança, a maioria dos respondentes (69%, $n = 100$) não ocupava cargos de gestão, enquanto 31% ($n = 45$) desempenhavam funções de liderança. Em relação ao tempo de atuação no órgão, 46% dos participantes ($n = 66$) indicaram ter mais de 15 anos de experiência na instituição. Outros 22% ($n = 32$) relataram estar na organização entre 0 a 3 anos, 14% ($n = 21$) entre 6 a 10 anos, 12% ($n = 17$) entre 11 a 13 anos, e 6% ($n = 9$) entre 4 a 5 anos. A variável idade revela que a maioria dos participantes tem entre 35 e 44 anos (37%, $n = 54$) ou entre 45 e 55 anos (30%, $n = 44$). Outros grupos etários incluem participantes de 55 a 64 anos (17%, $n = 25$), de 25 a 34 anos (14%, $n = 21$) e um respondente com 65 anos ou mais (1%, $n = 1$). Não houve participação de indivíduos com idade entre 18 e 24 anos.

Esses dados demográficos e profissionais oferecem um panorama da amostra, composta majoritariamente por servidores com significativa experiência na organização e com uma distribuição etária concentrada entre 35 e 55 anos.

3.3 Percurso Metodológico do Estudo 2

Teste de Hipóteses

Com o instrumento construído e validado, deu-se início ao Estudo 2, destinado ao teste das hipóteses relacionadas às variáveis individuais: Reação, Aprendizagem e Transferência. O instrumento e o procedimento de coleta de dados empregados foram os mesmos utilizados na subfase anterior para a AFC. Isto é: o questionário foi aplicado com servidores públicos de duas organizações distintas da Administração Pública federal que já haviam participado previamente do processo de avaliação de desempenho em seus respectivos órgãos, resultando em um total de 145 respostas. A característica da amostra de respondentes também é a mesma descrita na subseção anterior.

Para testar as hipóteses 1 e 2, aplicou-se a técnica de Regressão Linear Simples, com o objetivo de verificar a contribuição das variáveis independentes na variável dependente (Cozby, 2003; Hair Junior et al., 2005). A análise dos dados foi realizada no software JASP (versão 0.19.1), desenvolvido pela Universidade de Amsterdã, o qual permite realizar análises com *bootstrap*, o que complementa as estimativas obtidas dos testes estatísticos com estimativas de intervalos de confiança, não necessitando assumir que o intervalo de confiança do parâmetro segue uma distribuição normal (Haukoos & Lewis, 2005). Inicialmente, foram avaliadas as correlações entre as variáveis preditoras

(VI) e as variáveis critério (VD). Em seguida, apenas as relações que apresentaram significância estatística nas correlações foram submetidas ao teste de regressão linear.

4 RESULTADOS

Esta seção é composta pela apresentação e discussão dos resultados dos Estudos 1 e 2. O Estudo 1 está dividido nas Fase A e Fase B. A primeira fase está subdividida em duas subfases: na Subfase 1A, o foco está na construção das escalas, subsidiadas pelos dados qualitativos coletados nas entrevistas. Para isso, tem-se quatro processos: processamentos dos dados textuais e análise de conteúdo. Na subfase 2ª, são apresentados os resultados da validação de juízes.

Na Fase B, foram realizadas as análises para apresentação de evidências de validade. Na subfase 1B, os dados quantitativos foram inicialmente submetidos à análise fatorial exploratória (AFE), com o objetivo de identificar as dimensões subjacentes às escalas e verificar a estrutura dos fatores; e na Subfase 2B, foi realizada a análise fatorial confirmatória (AFC), adotada para testar e confirmar a estrutura de fatores identificada na etapa exploratória, assegurando a validade do modelo.

Após a construção e validação das escalas, foi realizado o Estudo 2, o qual se concentra no teste de hipóteses das relações entre as variáveis individuais. Nesta seção, as hipóteses foram testadas utilizando modelos estatísticos, com o objetivo de verificar as relações propostas entre as variáveis estudadas.

Resultados do Estudo 1

4.1. Fase A: Construção teórica e validação de juízes das Escalas de Percepção de Reação, de Aprendizagem e de Transferência

Esta seção busca demonstrar a construção e validação de juízes das medidas das dimensões Reação, Aprendizagem e Transferência. Para isso, uma primeira versão do instrumento de pesquisa se deu a partir de uma construção teórica decorrente de escalas identificadas na literatura revisada anteriormente. Esses instrumentos de referência possibilitaram a seleção isolada de itens ou de escalas inteiras com a posterior tradução e adaptação em um instrumento único de cada escala. Ao todo foram analisados mais de 200 itens entre as escalas revisadas para seleção dos itens pertinentes à teoria e à sua definição constitutiva, conforme tabela apresentada na seção de método.

Após a leitura, tradução e adaptação dos itens selecionados, três escalas foram inicialmente construídas: i) Escala de Percepção de Reação do Sistema de Gestão de

Desempenho (EPR-GD), com 106 itens; ii) Escala de Percepção de Aprendizagem do Sistema de Gestão de Desempenho (EPA-GD), com 36 itens; e iii) Escala de Percepção de Transferência do Sistema de Gestão de Desempenho (EPT-GD), com 66 itens.

Essas três versões iniciais das escalas compuseram o roteiro de entrevista aplicado aos servidores, sendo objeto central de análise da Seção 3, mas também permitiram preparar as perguntas das Seções 1 e 2. As entrevistas foram utilizadas para identificar os elementos constituintes da percepção da efetividade do sistema de GD a fim de verificar se e em que medida os critérios e dimensões selecionados anteriormente para construção das escalas eram abarcados na fala dos entrevistados. Isso também permitiu identificar outros aspectos que eventualmente pudessem ter sido negligenciados apenas pela construção da escala decorrentes da literatura. Os resultados das entrevistas foram descritos na subseção seguinte, chamada de Subfase 1A.

Subfase 1A: Construção das escalas – Resultados da Entrevistas

O conteúdo gerado a partir das entrevistas foi analisado com apoio do software *Iramuteq* e de análise de conteúdo de Bardin (2016). Para isso, inicialmente foi feita a transcrição do conteúdo das 11 entrevistas realizadas. O conteúdo foi codificado e organizado pela fala de cada um dos 14 entrevistados em um documento chamado *corpus* textual, necessário para análise no software *Iramuteq*.

O *corpus* textual contou com uma estrutura de 14 textos, com 27.148 ocorrências, entre palavras, formas ou vocábulos, sendo 3.500 palavras distintas e 1.821 com uma única ocorrência. Os textos foram separados em 776 segmentos de texto (STs) com aproveitamento de 644 (82,99%), sendo considerado um bom aproveitamento um índice de 75% ou mais (Camargo; Justo, 2018).

Nas subseções seguintes, são apresentados os resultados do processamento dos dados no software *Iramuteq*, que desempenhou um papel essencial na análise de conteúdo. A análise inicial focou na identificação de agrupamentos de palavras e conceitos relevantes, fornecendo subsídios para a construção das categorias. Esse processo possibilitou não apenas uma compreensão mais profunda das respostas, mas também contribuiu para integrar as categorias extraídas nas etapas prévias às entrevistas e incorporar aquelas desenvolvidas posteriormente, alinhando-as de forma coerente ao instrumento final. A análise textual, aliada às categorias identificadas, teve como objetivo estruturar e refinar os conteúdos das escalas de Reação, Aprendizagem e Transferência,

culminando em um instrumento mais robusto e alinhado às especificidades da Administração Pública brasileira.

4.1.1. Processamento dos Dados Textuais

Inicialmente, foi realizada uma análise gráfica de similitude que se baseia na teoria dos grafos, o que possibilita identificar as ocorrências entre as palavras e seu resultado traz indicações da conexão entre as palavras, auxiliando na identificação da estrutura analisada, contribuindo para verificar as variáveis a serem identificadas na análise de dados (Marchand; Ratinaud, 2012).

Na análise de similitude, foi possível identificar tanto a ocorrência das palavras quanto a conexão entre elas, conforme ilustrado na Figura 7. Os formatos coloridos representam o agrupamento de palavras que mais se relacionam, o que permitiu analisar os principais temas abordados pelos entrevistados. Com o objetivo de facilitar a leitura e visualização dos dados na figura, optou-se por incluir apenas as palavras que foram mencionadas pelo menos 40 vezes no *corpus* textual.

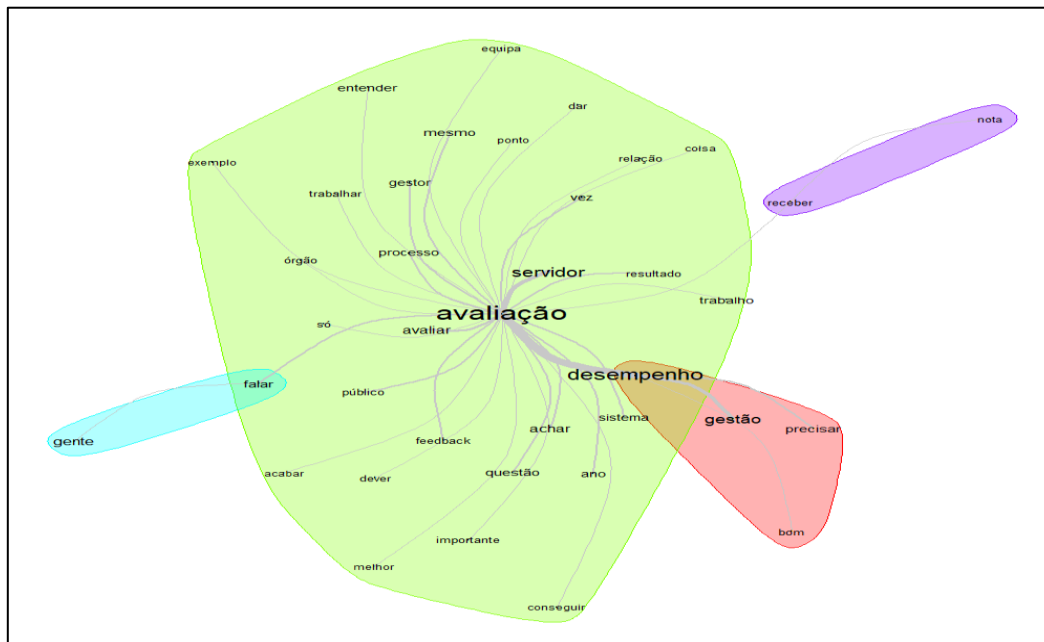


Figura 7. Ocorrência das palavras e conexão entre elas

Fonte: Elabora do pela autora.

A análise de similitude contemplou 32 palavras agrupadas em quatro comunidades (subgrupos de palavras) distintas. Quatro palavras se destacaram na fala dos

entrevistados abordaram aspectos mencionados pela literatura como cruciais para a efetividade do sistema de GD. No entanto, para uma compreensão mais aprofundada, é necessário explorar as estruturas lexicais para identificar quais critérios e dimensões foram contemplados nas falas dos entrevistados.

Para isso, foi utilizada a Classificação Hierárquica Descendente (CHD) a partir do software *Iramuteq*, a qual reflete a relação entre as palavras estatisticamente significativas. Nesse sentido, a CHD retornou quatro classes a partir do conteúdo das entrevistas, sendo que a Classe 1 contou com 204 STs (31,86%), a Classe 2, com 205 STs (31,83%), a Classe 3 contendo 107 STs (16,61%) e a Classe 4 contou com 128 STs (19,88%).

O software *Iramuteq* forneceu dendrograma (Figura 9) que possibilitou a representação das classes contendo as palavras com maior frequência média em relação umas às outras e com diferenças significativas entre elas.



Figura 9. Dendrograma das classes de palavras

Fonte: Elaborado pela autora.

As primeiras palavras de cada classe e que estão com a fonte em tamanho maior são aquelas que apresentam um maior chi-quadrado, o que reflete as palavras que mais se

associaram com as outras palavras da mesma classe, fornecendo indícios acerca da temática daquela classe. Para isso, recorreu-se aos segmentos de texto de cada palavra estatisticamente significativa de cada classe com base nos dados disponibilizados pelo software *Iramuteq*. Foi realizada uma leitura desses extratos, o que permitiu caracterizar tematicamente, por meio da interpretação pessoal da pesquisadora e sustentada na literatura, cada uma das classes. Uma breve descrição acerca dos aspectos identificados em cada uma das classes é apresentada a seguir, na Tabela 21.

Tabela 21

Descrição dos principais temas abordados por classes

Classe	Percentual do Corpus (%)	Palavras Principais (x ²)	Principais Temas
Classe 1	31,68%	Nota (59,99), Ciclo (42,62), Estratégico (25,81), Nível (25,63)	Desafios na prática de gestão de desempenho, percepção de uso mecânico da avaliação, falta de indicadores claros e alinhamento estratégico.
Classe 2	31,68%	Avaliação (53,45), Progressão (49,96), Carreira (40,22), Financeiro (19,15)	Desenho técnico-instrumental do sistema, relação entre a avaliação e aspectos financeiros, impactos negativos na efetividade do sistema.
Classe 3	16,61%	Perfil (50,98), Clareza (34,41), Objetivo (32,19), Peso (30,4)	Princípios desejáveis como transparência, basear-se em dados e customização, necessidade de feedback claro e desenvolvimento contínuo.
Classe 4	19,88%	Equipe (48,91), Experiência (43,02), Ferramenta (27,3), Comunicação (26,96)	GD como ferramenta organizacional, desafios técnicos e de comunicação, importância do engajamento e desenvolvimento de líderes, desafios da avaliação no trabalho híbrido.

Conclui-se que as categorizações geradas pelo processamento de dados no software *Iramuteq*, organizadas em quatro classes, ofereceram indicadores sobre a percepção da efetividade do sistema de GD. De modo geral, os entrevistados relataram desafios significativos na condução do sistema, os quais dificultam o alcance dos resultados esperados, evidenciando componentes essenciais para a efetividade do sistema de GD.

Esses temas oferecem subsídios relevantes para a análise de conteúdo subsequente, mostrando convergências com as categorias inicialmente propostas. Entre elas, destacam-se: **atitudes no trabalho**, como o engajamento de líderes (Classe 4);

habilidades, relacionadas ao desenvolvimento de competências de líderes para realizar avaliações precisas (Classe 3 e 4); e **utilidade**, enfatizando a necessidade de um desenho do sistema menos técnico-instrumental e mais estratégico, que se posicione como uma ferramenta organizacional robusta (Classe 2). Esses e outros aspectos serão explorados em maior detalhe na seção seguinte, aprofundando as reflexões sobre a efetividade do sistema de GD. Esse processo visa validar e complementar, se necessário, o conteúdo das escalas inicialmente propostas, permitindo a identificação de aspectos que não foram abordados pelos entrevistados ou a incorporação daqueles que não haviam sido contemplados inicialmente.

4.1.2. Análise de Conteúdo dos Dados Semiestruturados

Um dos principais objetivos das entrevistas consistiu em verificar a adaptação das categorias inicialmente propostas e dos itens traduzidos de escalas internacionais, próprias de outros contextos, ao contexto da Administração Pública brasileira. Assim, esta subseção reúne as percepções dos 14 entrevistados sobre aspectos importantes que caracterizam um sistema de GD em suas organizações, evidenciando os desafios enfrentados e destacando os aspectos que distanciam o sistema dos resultados esperados.

Assim, o roteiro de entrevista que continha questões semiestruturadas, nas seções 1 e 2, gerou conteúdo que foi distribuído nas próprias categorias que compõem o modelo teórico-integrativo deste estudo, com foco nas variáveis individuais a serem testadas, cujas escalas estavam em construção: Reação, Aprendizagem e Transferência. Ao todo, 13 critérios, divididos entre essas três dimensões, fizeram parte da análise de conteúdo das entrevistas, não tendo sido identificada a necessidade de criação ou aglutinação de outras categorias de análise.

Nos conteúdos seguintes, apresenta-se cada uma das categorias que foi analisada e mencionada pelos entrevistados com base nas perguntas semiestruturadas (Seção 1 e Seção 2). Os resultados podem ser conferidos na Tabela 22.

Tabela 22

Categorias de efetividade do sistema de GD relatadas nas entrevistas

Categoria	Subcategoria	Recorrência de Relatos		Frequência de Indivíduos
Reação	Utilidade	9		6
	Precisão	8	17	5
	Satisfação	-		-

	Justiça	-	-
	Afetividade	-	-
Aprendizagem	Conhecimento	6	5
	Atitude	4	12
	Habilidade	2	2
Transferência	Relação interpessoal	3	2
	Desempenho no trabalho	2	2
	Motivação	1	5
	Atitudes no trabalho	-	-
	Bem-estar	-	-

Nota: As categorias sem recorrência de relatos ou frequência entre os indivíduos não foram mencionadas nas seções 1 e 2 do roteiro de entrevista.

De modo geral, o conteúdo das entrevistas apresentou maior recorrência na dimensão Reação ($n = 17$), seguido de Aprendizagem ($n = 12$), e, por último, a Transferência ($n = 5$). A categoria de Aprendizagem contou com seus três critérios amplamente mencionados pelos entrevistados durante as seções semiestruturadas do roteiro. Os participantes destacaram a importância de um sistema de GD efetivo em fornecer clareza sobre os comportamentos esperados dos colaboradores e apontaram a necessidade de desenvolver habilidades tanto para fornecer quanto para receber feedback de maneira adequada. Esse achado contrasta em partes com a literatura teórica, em que a dimensão Aprendizagem é relativamente mais recente e menos consolidada pelos autores.

Já a dimensão Reação e a dimensão Transferência tiveram critérios que não foram explicitamente abordados pelos entrevistados durante o conteúdo da seção semiestruturada do roteiro. Contudo, essas categorias ganharam relevância na seção estruturada do roteiro (Seção 3), quando as escalas contendo as dimensões, critérios e definições foram apresentadas. Nessa etapa, os entrevistados, de modo geral, corroboraram a categorização proposta e reforçaram a necessidade dessas dimensões para compreender a efetividade do sistema de GD.

A única exceção foi a categoria Satisfação, da dimensão Reação, que gerou desconforto entre os entrevistados devido à percepção de duplicidade de informações. Esse aspecto resultou na exclusão da categoria, conforme descrito mais detalhadamente nesta seção de resultados. Essa exclusão reflete principalmente a preocupação em manter a coesão e a clareza da estrutura das escalas.

Nos trechos a seguir, é possível conferir as categorias analisadas e os respectivos trechos dos entrevistados que as corroboram.

Categoria Reação

A dimensão Reação foi mencionada pelos entrevistados principalmente nas categorias Precisão e Utilidade, evidenciando alinhamento com os itens originalmente propostos para a escala de Reação, conforme achados resumidos na Tabela 23. A análise não identificou a necessidade de criar novas categorias ou reformular os critérios existentes. Embora as demais categorias não tenham sido diretamente abordadas pelos entrevistados nas perguntas semiestruturadas, os resultados obtidos na seção estruturada do roteiro demonstraram adequação e consistência, validando especialmente a inclusão dos critérios de Justiça e Afetividade. No entanto, o critério de Satisfação apresentou um comportamento distinto, gerando percepções diferenciadas entre os entrevistados, cuja análise será aprofundada em seção posterior.

Tabela 23

Principais achados sobre Reação apontado pelas entrevistas

Categoria	Subcategoria	Principais Achados das Entrevistas
Reação	Utilidade	- Alinhamento do sistema de GD com objetivos organizacionais, promovendo desenvolvimento profissional além da gratificação financeira. - Percepção de profundidade e relevância nas avaliações, impulsionando o crescimento e aprimoramento contínuo dos servidores. - Utilização do sistema de GD como ferramenta estratégica para movimentação funcional e alocação eficiente de talentos, otimizando os recursos humanos da organização.
	Precisão	- Capacidade do sistema de GD de diferenciar adequadamente os desempenhos para identificar lacunas e promover o desenvolvimento profissional. - Uso de indicadores claros e objetivos que garantam avaliações consistentes e imparciais. - Disponibilidade de feedback assertivo e contínuo ao longo do ciclo avaliativo, proporcionando oportunidades de melhoria. - Adequação e atualização regular dos critérios e métricas para refletir com precisão o desempenho real dos colaboradores.

A subcategoria identificada foi a **Utilidade**, referente à capacidade do sistema de GD em fornecer insumos para outras políticas e práticas da organização (Ikramullah et al., 2016). Tal como apontado por Keeping e Levy (2000), esse critério é um dos mais

consistentes na literatura e, pela quantidade de relatos, um dos que os entrevistados mais identificam.

Nesse sentido, destaca-se a importância conferida pelos entrevistados ao alinhamento do sistema de GD com objetivos organizacionais, promovendo desenvolvimento profissional, além da gratificação financeira. Esse aspecto é representado pela maior crítica dos entrevistados quanto à vinculação de GD à gratificação financeira, sendo apontada como uma das principais causas de distorções no processo avaliativo. Essa conexão transforma o que deveria ser uma ferramenta de desenvolvimento profissional e melhoria dos serviços prestados em uma formalidade, cujo objetivo maior é garantir o aumento da remuneração dos servidores.

“Isso faz com que as pessoas tenham uma cultura em que acreditam que seu desempenho precisa estar atrelado ao financeiro e não ao bem maior, que é o resultado e a entrega ao serviço público e à sociedade em geral” (Indivíduo 6).

“Em relação a essas questões, acho que estamos falhando miseravelmente em avaliar desempenhos e em fazer gestão dos desempenhos em prol do serviço público e das entregas que a administração pública precisa e deve realizar. [...] A avaliação de desempenho tal como foi concebida até hoje não dá conta” (Indivíduo 14).

Outro ponto destacado pelos entrevistados é que essa ligação entre desempenho e gratificação financeira gera uma percepção de que o objetivo final da avaliação é o financeiro, e não o desenvolvimento e o aprimoramento de comportamentos no trabalho.

“A progressão funcional não deveria ter relação com o financeiro, mas sim com um impacto indireto, e não direto, como o resultado da avaliação de desempenho. [...] Para mim faz mais sentido se fosse feita uma avaliação real do desempenho do servidor, que poderia ser utilizada para promoções ou ocupações de gratificações” (Indivíduo 13).

Essa visão reforça a necessidade de repensar o sistema de GD para que ele se afaste do mero viés financeiro e passe a cumprir seu papel central de promover o crescimento profissional e a melhoria do serviço público. A falta de uma “avaliação real” também foi citada por outros entrevistados, que ressaltam que, embora o processo tenha passado por algumas melhorias ao longo dos anos, ele ainda está muito atrelado à questão da remuneração.

A análise sugere que o foco excessivo na gratificação financeira desvia a missão do serviço público, substituindo a busca pelo desenvolvimento contínuo dos servidores pela obtenção de benefícios financeiros. Ao invés de incentivar o crescimento profissional

e a entrega de melhores resultados para a sociedade, o sistema de GD parece promover uma cultura de superficialidade, na qual a avaliação se resume a um mecanismo de progressão automática, sem uma análise profunda e realista do desempenho.

A superficialidade do processo avaliativo é percebida como um dos maiores problemas, sendo mencionada em diversos relatos. Por exemplo, o Indivíduo 1 descreve um cenário em que praticamente todos os servidores recebiam notas máximas, o que tornava impossível diferenciar o bom do excelente desempenho, impossibilitando ações como uma progressão de carreira.

“O negócio era completamente proforma e todo mundo praticamente tirava 10. Então não era possível distinguir o bom desempenho do excelente desempenho para que uma pessoa recebesse uma avaliação de desempenho que afetasse sua progressão na carreira” (Indivíduo 1).

Essa visão é compartilhada por outros entrevistados, que apontam que a avaliação de desempenho na Administração Pública se tornou um processo automático, no qual os mesmos indicadores são utilizados repetidamente, sem considerar o impacto real das atividades realizadas pelos servidores. Além disso, as pressões políticas e institucionais também influenciam negativamente a integridade do processo avaliativo. O Indivíduo 8, por exemplo, menciona que, em muitos casos, as notas são “arredondadas” para que os servidores não sofram prejuízos financeiros:

“Arredondamos o resultado para que o servidor não tenha prejuízo financeiro. E se houver discordância, o servidor pode entrar com recurso contra o chefe, no entanto, há questões políticas envolvidas e o chefe muitas vezes acaba sendo refém dos serviços públicos” (Indivíduo 8).

Essa prática de arredondamento de notas, juntamente da influência política sobre os gestores, enfraquece o sistema de GD, transformando-o em um processo pouco eficaz para identificar e corrigir deficiências no desempenho dos servidores. Além disso, muitos entrevistados apontam que, em algumas carreiras, a avaliação de desempenho é o único critério para a progressão, o que gera distorções ainda maiores. Esses fatores evidenciam uma forte dependência da gratificação financeira, o que dificulta o uso da GD como uma ferramenta de melhoria contínua. Como aponta o Indivíduo 13, essa situação é particularmente prejudicial em algumas carreiras públicas.

“A avaliação de desempenho está atrelada à questão da gratificação, o que gerou uma estagnação na maioria das carreiras que não possuem progressão e

dependem apenas desse processo avaliativo vinculado ao financeiro. [...] Há grandes questões em relação à progressão funcional, que tem como resultado a avaliação de desempenho para algumas carreiras, mas não para outras, e isso seria um ponto positivo” (Indivíduo 13).

Em suma, as entrevistas sugerem que a vinculação da GD à remuneração tem gerado uma série de efeitos adversos, como a superficialidade nas avaliações e a perda de foco no desenvolvimento profissional. Para que o sistema de GD cumpra sua função primordial de promover o desenvolvimento dos servidores e a melhoria dos serviços prestados à sociedade, torna-se necessário rever a relação entre os propósitos do sistema de GD e a sua relação com outros aspectos que podem vir a impedir a utilização adequada do sistema de GD. Essas percepções de que as avaliações são, muitas vezes, formais e pouco fidedignas, representando instrumentos *proforma*, não figurava entre os itens inicialmente traduzidos e adaptados aos instrumentos das escalas deste estudo. Sendo assim, esse foi um dos conteúdos que forneceram subsídios para inclusão de um novo item relacionado ao uso do sistema de GD apenas *proforma*.

Tendo em vista que a utilidade se refere ao quanto a ferramenta atende seu propósito (*e.g.* Abdullah & Abraham, 2012), destaca-se também a efetividade do sistema de GD poder ser aferida também pela sua utilização como ferramenta estratégica para movimentação funcional e alocação eficiente de talentos, otimizando os recursos humanos da organização. Os entrevistados enfatizaram a importância de que o sistema de GD seja genuíno, isto é, que vá além da mera formalidade e busque, de fato, identificar o desempenho real dos servidores para que a avaliação sirva como base não apenas para progressão na carreira, mas também para a promoção e ocupação de cargos, facilitando a movimentação dentro da organização e melhorando a alocação de talentos.

“Se houvesse uma avaliação de desempenho, talvez seria possível identificar melhor qual pessoa está mais adequada para assumir determinada função ou órgão, mas como não há, é tudo na loteria. É uma grande mudança na administração pública” (Indivíduo 10).

A falta de um sistema robusto de avaliação de desempenho impacta diretamente a capacidade de identificar servidores aptos para novas funções, criando uma sensação de aleatoriedade nas decisões de movimentação interna. Isso é particularmente problemático na Administração Pública, onde as trajetórias profissionais são diversas e os processos seletivos via concurso público não abordam questões comportamentais ou competências específicas para cargos de liderança.

“Se o servidor tem o desejo de se movimentar dentro da minha organização, dentro da administração pública, o resultado da avaliação de desempenho poderia ser usado como um critério positivo a meu favor” (Indivíduo 13).

“À medida que o concurso público não dá conta de algumas questões, você terá que fazer uma gestão da porta para dentro” (Indivíduo 14).

Essas falas destacam a necessidade de um sistema de avaliação que vá além das qualificações técnicas, incluindo a identificação de competências comportamentais e a análise de adequação para diferentes funções e contextos organizacionais. A percepção dos entrevistados sobre a utilidade do sistema de GD evidencia uma convergência com a literatura e reforça a pertinência dos itens do questionário utilizado nesta pesquisa. Os relatos destacam que, embora o sistema de GD tenha potencial para subsidiar políticas de movimentação funcional, progressão de carreira e desenvolvimento profissional, sua aplicação atual, especialmente em organizações públicas, tem se mostrado limitada e distorcida. A vinculação excessiva à gratificação financeira emerge como o principal obstáculo, desviando o foco de práticas que poderiam aprimorar a gestão de talentos e promover o alinhamento entre competências individuais e demandas organizacionais.

A análise indica que os itens do questionário que tratam da Utilidade do sistema de GD devem capturar os aspectos destacados pelos entrevistados, fornecendo insumos confiáveis para decisões relacionadas à movimentação interna e promoção de servidores. Os relatos dos participantes reforçam a necessidade de uma escala que contemple os elementos fundamentais discutidos na literatura, como apontado por Ikramullah et al. (2016) e Keeping e Levy (2000).

A outra subcategoria identificada foi **Precisão**, que contempla a capacidade do sistema de GD de refletir corretamente aquilo a que ele se propõe (Sharma et al., 2016). O primeiro ponto refere-se à importância de o sistema de GD ser capaz de diferenciar adequadamente os desempenhos para identificar lacunas e promover o desenvolvimento profissional. Os entrevistados destacaram que a disparidade entre a nota atribuída e o real desempenho avaliado compromete a efetividade do sistema. Relatos frequentes apontaram uma tendência generalizada de atribuir notas máximas, independentemente do desempenho, o que distorce o propósito central da avaliação: identificar lacunas e promover o crescimento profissional.

“É estranho considerar que todos estão ali recebendo uma nota muito próxima da nota máxima. Se todo mundo está ótimo num nível excepcional, o que vou

desenvolver? Será que é isso mesmo que estamos entregando para a sociedade?” (Indivíduo 11).

“Para ter uma ideia: nos últimos 5 anos, a média de nota máxima foi em torno de 96 na instituição. Então você quer dizer que só temos craques, certo?” (Indivíduo 8).

“É um aspecto que parece que a gente tem todos os empregados trabalhando na NASA, porque a média das notas é acima de 95, onde o teto é 100” (Indivíduo 12).

“Todo mundo tem nota máxima, então, não, ele [sistema de gestão de desempenho] não subsidia outras ações muito difíceis, muito difíceis. É o que temos. Acho que a parte de desenvolvimento é muito agradável de trabalhar, não é o desempenho” (Indivíduo 11).

Esses comentários refletem como essa prática prejudica a integridade do sistema e destaca a necessidade de maior honestidade e objetividade nas avaliações. Além disso, a ausência de indicadores claros e objetivos, conforme relatado, gera subjetividade e erros de julgamento, afetando a confiança dos servidores no sistema. A precisão na atribuição de notas é apontada como fundamental para garantir que o sistema de GD seja adequado e eficaz, como descrito por Sharma et al. (2016). Para os autores, um sistema de GD deve ser capaz de relacionar os desempenhos individuais apresentados baseados naquilo que foi planejado, indicando por meio de feedbacks e aconselhamentos regulares.

“É importante [...] que esses indicadores sejam medidos e monitorados constantemente” (Indivíduo 12).

Essa abordagem baseada em dados objetivos e precisos mostra-se essencial para superar práticas avaliativas mecânicas e proforma, que apenas reproduzem processos sem impactar de forma significativa o desenvolvimento individual e organizacional.

A acurácia do processo de GD também é comprometida pela falta de comunicação clara sobre os critérios e expectativas. Embora os critérios existam formalmente, como mencionado pelo Indivíduo 14, eles não são aplicados de forma clara e transparente na prática. Isso resulta em um desalinhamento entre o que os colaboradores entendem que é esperado deles e o que realmente é avaliado, criando confusão e frustrações durante o ciclo avaliativo. Tal situação reforça a percepção de que o sistema de GD não é preciso na identificação e medição do desempenho dos colaboradores, o que pode afetar sua credibilidade.

Outro aspecto identificado é com relação ao uso de indicadores claros e objetivos que garantam avaliações consistentes e imparciais. Os relatos sugerem que a ausência

desses indicadores claros e precisos faz com que a avaliação seja subjetiva, reforçando percepções de injustiça. A falta de precisão na avaliação é evidenciada no relato do entrevistado, que questiona a validade da nota atribuída em relação à assiduidade, mesmo sem ter faltado ao trabalho. Esse exemplo revela como a imprecisão dos indicadores avaliados e a ausência de critérios objetivos levam a erros de julgamento, prejudicando a confiança no sistema de GD, tal como apontado por Iqbal et al. (2015), como erros e vieses que acometem o sistema de GD.

“Eu, por exemplo, recebi uma nota menor por conta da assiduidade, mesmo nunca tendo faltado e questioneei isso” (Indivíduo 7).

Outro ponto enfatizado foi a disponibilidade de feedback assertivo, elemento essencial para garantir que o sistema seja percebido como preciso (Sharma et al., 2016). A falta de feedback assertivo e contínuo ao longo do ciclo avaliativo contribui para a desconexão entre o que se espera dos colaboradores e o que é efetivamente avaliado, dificultando correções e melhorias. Além disso, o entrevistado chama de falta de uma “conversa estimulante” durante a avaliação.

“[...] O problema é que não há uma clareza a respeito do que é esperado das pessoas e não há uma conversa estimulante na hora da avaliação” (Indivíduo 7).

Por último, outro aspecto consiste na adequação e atualização regular dos critérios e métricas para refletir com precisão o desempenho real dos colaboradores. Iqbal et al. (2015) também apontam a necessidade de considerar os procedimentos do sistema de GD, o qual pode interferir na forma com que as pessoas concordam ou buscam burlar regras da política de GD. Os entrevistados reforçaram essa necessidade de ajustes no desenho do sistema, incluindo o uso de métricas e indicadores de desempenho objetivos e precisos. Segundo o entrevistado, um ajuste no desenho do sistema é importante para ajudar a evitar distorções como a atribuição sistemática de notas máximas na avaliação.

“Estamos colhendo informações do nosso piloto que foi aplicado no passado e o que estamos percebendo é que estamos conseguindo identificar quem são as pessoas de alto potencial. [...] Já não temos mais todo mundo ganhando 100 pontos” (Indivíduo 2).

Portanto, os conteúdos relatados pelos entrevistados estão em forte consonância com os aspectos apontados pela literatura sobre a precisão em sistemas de GD. Essa

convergência entre teoria e prática oferece subsídios importantes para justificar os itens relacionados à precisão.

Categoria Aprendizagem

A dimensão Aprendizagem foi mencionada pelos entrevistados, abrangendo todos os três critérios originalmente propostos: Conhecimento, Habilidade e Atitude. A análise não identificou a necessidade de criar novas categorias ou reformular os critérios existentes, indicando que os itens propostos atendem às percepções relatadas. Os principais aspectos apontados pelos entrevistados de cada uma das categorias estão reunidos na Tabela 24.

Tabela 24

Principais achados sobre Aprendizagem apontados pelas entrevistas

Categoria	Subcategoria	Principais Achados das Entrevistas
Aprendizagem	Conhecimento	- Difusão sobre as particularidades de cada função e unidade, considerando competências técnicas e comportamentais. - Clareza das expectativas é crucial para engajamento e desempenho. - Transparência sobre os comportamentos mais estratégicos a partir do alinhamento estratégico entre GD e metas organizacionais para evitar avaliações fragmentadas e descontextualizadas. - Informações alinhadas entre níveis estratégico e operacional para o cumprimento de objetivos organizacionais e o desenvolvimento individual.
	Atitude	- Aversão à participação no GD por parte de servidores e gestores, frequentemente motivada pela cultura organizacional e pela resistência a feedbacks críticos. - Resistência mais acentuada entre servidores antigos, que veem o processo com desconfiança e hesitam em participar de avaliações voltadas ao progresso. - Sensibilização e treinamento dos líderes favorecem um ambiente mais participativo e propenso ao desenvolvimento contínuo.
	Habilidade	- Desenvolvimento de líderes para fornecer feedback construtivo e conduzir avaliações de forma adequada, promovendo aprendizado contínuo. - Preparo dos gestores com formações específicas para avaliadores.

O critério **Conhecimento** é relacionado às informações declarativas, procedimentais ou tácitas adquiridas em decorrência da interação com o sistema de GD

(Schleicher et al., 2019). Os entrevistados enfatizaram a necessidade de adaptar o sistema de GD às particularidades de cada função e unidade. Para que o sistema seja eficaz, destacou-se que ele deve considerar tanto as competências técnicas quanto as comportamentais, além de fornecer clareza sobre o que é esperado de cada servidor.

“Podemos definir quais competências serão avaliadas para cada função específica, bem como o peso do trabalho técnico da entrega e do perfil comportamental. [...] Nossa intenção é possibilitar a customização da avaliação de desempenho para cada função com base em alguns parâmetros” (Indivíduo 1).

A personalização do processo avaliativo com base nas exigências específicas de cada função torna o sistema de GD mais relevante e alinhado às necessidades organizacionais, facilitando a identificação de áreas em que o servidor precisa se desenvolver.

“Podemos identificar quais competências, habilidades e conhecimentos aquele servidor precisa para exercer aquela função” (Indivíduo 5).

Esse achado é consistente com a literatura, que aponta a clareza das expectativas como um elemento crucial para que os funcionários compreendam suas responsabilidades e tenham maior propensão a alcançar as metas estabelecidas (*e.g.*, Ikramullah et al., 2016; Kang & Choi, 2019; Teeroovengadum et al., 2019). Vários participantes mencionam a importância de se ter essa clareza sobre as expectativas e os critérios de avaliação para evitar situações ou cobranças inesperadas. A clareza no processo de GD é fundamental para garantir que os colaboradores saibam exatamente o que precisam fazer para alcançar o sucesso na organização. O Indivíduo 1 enfatiza a importância de ter clareza sobre seu papel, as expectativas e os caminhos para o desenvolvimento e crescimento na organização. A falta dessa clareza limita a capacidade dos funcionários de se prepararem e se desenvolverem de acordo com as metas organizacionais, contribuindo para um sentimento de estagnação e falta de direcionamento.

“Os critérios estão postos, mas não há clareza a respeito daquilo que se espera na prática” (Indivíduo 14).

“Percepção de ter uma chance de crescer chance de se desenvolver, clareza do meu papel, clareza do que é esperado de mim, clareza do que preciso fazer para ter sucesso na organização” (Indivíduo 1).

Para que o sistema de GD funcione adequadamente, é essencial que os critérios avaliativos sejam aplicados de forma clara e comunicados de maneira transparente, permitindo que os colaboradores alinhem suas ações às demandas organizacionais. Segundo Abdullah e Abraham (2012), maior clareza no processo de GD aumenta a probabilidade de os colaboradores adotarem uma atitude participativa em relação ao sistema.

Outro aspecto relevante destacado foi a necessidade de alinhamento estratégico vertical entre as políticas e práticas de GD e a estratégia organizacional. A ausência de um direcionamento claro e metas institucionais específicas dificulta a eficácia do processo avaliativo, tornando-o fragmentado e descontextualizado. Essa lacuna foi relatada por um entrevistado. O Indivíduo 8 expressa essa dificuldade ao mencionar que, em sua organização, as metas institucionais não incluem a área de gestão de pessoas, criando um sentimento de desalinhamento e falta de pertencimento.

“Eu não me enxergo em lugar nenhum, não tenho uma meta imediata”
(Indivíduo 8).

Sem um alinhamento estratégico adequado, os colaboradores podem ficar sem um direcionamento objetivo para guiar suas atividades, recorrendo a documentos antigos, como o regimento interno, que podem não estar mais em sintonia com a realidade atual da organização.

A falta de alinhamento estratégico também afeta a coesão entre os níveis estratégico e operacional. Isso foi particularmente enfatizado no contexto de ministérios, onde o desdobramento do Plano Plurianual (PPA) para objetivos setoriais e individuais é visto como um grande desafio.

“No nível operacional, ficamos aquém do nível estratégico, especialmente quando se trata de ministérios. Acredito que precisamos pensar muito nesse desdobramento começando pelo PPA (Plano Plurianual), que é feito para [durar] quatro anos ... depois disso é preciso planejar cada organização setorial e definir os objetivos internos até o nível dos indivíduos. Acho que esse é um grande desafio: fazer esse alinhamento. Para que dê certo é importante investir muito nessa fase de planejamento” (Indivíduo 11).

A análise reforça que, sem um alinhamento estratégico adequado, o sistema de GD carece de coesão, prejudicando tanto o desenvolvimento individual quanto o progresso organizacional. Essa desconexão impede que os servidores tenham clareza

sobre os comportamentos esperados e limita a eficácia do sistema em alcançar seus objetivos.

Assim, os relatos dos entrevistados sublinham a importância de clareza das atribuições e expectativas para o sucesso do sistema de GD. A ausência dessa clareza compromete sua capacidade de promover desempenho e engajamento, destacando a necessidade de investimento em comunicação e transparência dos critérios avaliativos. Dessa forma, os itens dispostos na escala de Aprendizagem convergiram com os resultados, pois abordavam o quanto o sistema de GD permitiu obter mais informações sobre as responsabilidades, tarefas e expectativas.

A segunda categoria **Atitude** refere-se à intenção de se comportar ou agir em decorrência da interação com o sistema de GD (Schleicher et al., 2019). Durante as entrevistas, emergiu como um aspecto relevante a aversão à participação no processo de GD, refletindo barreiras culturais e atitudinais que comprometem a eficácia do sistema.

O Indivíduo 13 destacou que muitos ocupantes de cargos de chefia estão habituados a receber notas máximas, o que gera uma sensação de intocabilidade em relação ao seu desempenho. Quando gestores tentam oferecer feedback que resulta em leves reduções nas notas, as reações dos servidores são frequentemente negativas, a ponto de alguns recorrerem formalmente contra a avaliação. Essa situação reflete a ausência de preparo e treinamento adequados para lidar com feedbacks construtivos.

“Eles [ocupantes de cargo de chefia] estão acostumados a sempre receber a nota máxima. Quando eu tenho um gestor minimamente preparado ele aponta um item, por exemplo, de gestão de equipe. Ele tira 0,25 pontos e esse servidor não aceita. Esse servidor entrou com um pedido de recurso porque ele não estava preparado para receber a avaliação. Ele não tinha tido nenhum treinamento. Ele recebeu uma nota com 0,25 a menos do resultado geral. Isso é cultural” (Indivíduo 13).

A resistência à avaliação, tanto por parte de servidores quanto de gestores, surge como um obstáculo significativo para a implementação de processos mais justos e transparentes. Essa resistência está profundamente enraizada na cultura organizacional, como relatado pelo Indivíduo 11.

“Não é apenas no serviço público. Eu acredito que na iniciativa privada também seja um pouco [difícil], porque é uma questão cultural. Hoje não fazemos uma avaliação real, é apenas uma avaliação” (Indivíduo 11).

No setor público, essa resistência é apontada como sendo mais acentuada entre servidores mais antigos, que já passaram por diversas progressões ao longo de suas carreiras. Muitos desses servidores veem a avaliação de desempenho com desconfiança e resistência, como menciona o Indivíduo 5.

“Os servidores mais antigos, especialmente aqueles que já tiveram 10 ou 12 progressões, não gostam muito da ideia de serem avaliados para progredir. Ainda há um momento bastante turbulento em relação à progressão e avaliação de desempenho aqui” (Indivíduo 5).

Esses relatos refletem a necessidade de uma mudança cultural mais ampla no serviço público, incluindo treinamentos específicos e uma reavaliação das práticas institucionais que perpetuam a superficialidade no processo avaliativo. O Indivíduo 4 ressaltou que, com o sistema de GD adequado, líderes conseguem perceber melhor as necessidades de desenvolvimento, demonstrando como o desenho e os procedimentos do sistema podem criar condições mais favoráveis.

“Aqueles que participaram do primeiro [ciclo de avaliação das lideranças] tiveram uma maior sensibilização em relação à necessidade de seu desenvolvimento como líderes. [...] Construir um sistema que permita que esses espaços de melhoria sejam registrados e sejam objeto de diálogo, sem muito receio” (Indivíduo 4).

A análise aponta que o engajamento com o sistema de GD está diretamente relacionado à forma como o processo é culturalmente percebido e operacionalizado. A aversão à participação, frequentemente motivada pela falta de preparo para receber feedbacks e pela resistência a avaliações críticas, não se limita a uma questão cultural, mas evidencia também a necessidade de sistemas mais bem desenhados e práticas institucionais que promovam confiança e transparência.

Além disso, o papel dos líderes se destaca como fator central para aumentar a adesão ao sistema. A sensibilização e o treinamento adequados dos gestores podem criar um ambiente mais propício ao desenvolvimento contínuo e à adoção de posturas participativas. Assim, os aspectos mencionados pelos entrevistados encontram respaldo nos itens relacionados à Atitude na escala de Aprendizagem.

O último critério é a **Habilidade**, relacionado à capacidade de aplicar os conhecimentos adquiridos com o sistema de GD (Ikramullah et al., 2016; Lawler, 2003; Schleicher et al., 2019). Os entrevistados destacaram a importância de desenvolver as habilidades dos servidores, especialmente dos gestores, para que possam conduzir

adequadamente o processo de avaliação. A falta de preparo de alguns líderes foi apontada como um obstáculo significativo à implementação bem-sucedida do sistema de GD.

“Dependendo do perfil da chefia, não se pode questionar muito, e há esse despreparo também dos líderes. [...] Agora há uma tendência de melhora com muitos cursos preparando lideranças, mas ainda há muito a ser feito” (Indivíduo 7).

O desenvolvimento contínuo de líderes é essencial para garantir que eles forneçam feedback construtivo e gerenciem o processo avaliativo de maneira eficaz, promovendo um ambiente de aprendizado contínuo. As habilidades desenvolvidas pelos gestores enquanto avaliadores foram identificadas na literatura como um dos resultados mais relevantes para um sistema de GD efetivo (*e.g.*, Haines & St-Onge, 2012; Rolle & Klinger, 2012; Teeroovengadum et al., 2019; Van Waeyenberg & Decramer, 2018; Yang & Hsieh, 2007).

“Eu não posso chegar no último dia do ano para avaliar e dar uma nota baixa, isso precisa ser um processo. É esse preparo que eu preciso para esse processo” (Indivíduo 13).

Conforme apontado por Haines e St-Onge (2012), o desenvolvimento de habilidades específicas pelos gestores, relacionado ao papel que desempenham no processo de GD, é um fator determinante para garantir que os resultados do sistema sejam atingidos. Esses autores enfatizam que sistemas de GD são percebidos como mais eficazes quando os avaliadores recebem uma formação mais extensa e aprofundada, capacitando-os a desempenhar suas funções de forma mais assertiva.

A análise reforça que o investimento no desenvolvimento das habilidades dos gestores é fundamental para a eficácia do sistema de GD, convergindo com a categoria e os itens inicialmente propostos. Líderes bem preparados não apenas tornam o processo avaliativo mais justo e construtivo, mas também contribuem para criar uma cultura organizacional voltada para o aprendizado e o aprimoramento contínuos. Esses elementos são apontados como cruciais para que o sistema de GD alcance seus objetivos e promova o alinhamento entre os resultados individuais e organizacionais.

Por fim, os três critérios da dimensão Aprendizagem – Conhecimento, Habilidade e Atitude – foram considerados elementos fundamentais pelos entrevistados para a efetividade do sistema de GD. A análise não indicou a necessidade de criar novas

categorias ou reformular os critérios existentes, confirmando que a estrutura atual é consistente, abrangente e alinhada com os desafios relatados pelos participantes.

Categoria Transferência

A dimensão Transferência foi amplamente mencionada pelos entrevistados, com destaque para as categorias Desempenho no Trabalho, Motivação e Relações Interpessoais, o que evidencia um alinhamento consistente com as categorias originalmente propostas para a escala de Transferência. Esses achados estão resumidos na Tabela 25. A análise das entrevistas não indicou a necessidade de criar novas categorias ou reformular os critérios existentes, demonstrando que os itens propostos para essa dimensão capturam adequadamente as percepções dos participantes.

Tabela 25

Principais achados sobre Transferência apontados pelas entrevistas

Categoria	Subcategoria	Principais Achados das Entrevistas
Transferência	Relação Interpessoal	- Feedback consistente e construtivo foi destacado como central para fortalecer a relação entre líderes e liderados. - Sistemas de GD bem implementados aumenta a receptividade e o engajamento dos colaboradores com o processo avaliativo. - Relações pautadas na realização de feedback contínuo durante o ciclo avaliativo com foco no desenvolvimento efetivo dos servidores.
	Desempenho no trabalho	- Avaliação de desempenho foi destacada como ferramenta para promover resultados organizacionais e eficiência. - Mecanismo para gestão de talentos e alocação estratégica de servidores, maximizando suas contribuições.
	Motivação	- Oportunidades de crescimento pessoal promovem melhorias nos comportamentos individuais. - Avaliações honestas e construtivas são apontadas como aspectos motivadores para os servidores.

O critério **Relação interpessoal** refere-se à interação social entre os indivíduos da organização (Schleicher et al., 2019). Os entrevistados destacaram que um sistema de GD bem implementado pode fortalecer as relações entre líderes e liderados, promovendo um

ambiente mais receptivo e colaborativo. Nesse contexto, a atribuição de feedback foi apontada como um mecanismo central para aumentar o engajamento dos servidores quando realizado de maneira consistente e construtiva.

“Já passamos pelo primeiro ano de avaliação de desempenho para a progressão, e hoje já está até melhor. As pessoas sentem que os gestores estão mais receptivos, escutando mais e [...] elas recebem um feedback do gestor que antes não recebiam” (Indivíduo 5).

A criação de uma cultura de feedback regular não apenas fortalece as relações interpessoais, mas também contribui significativamente para o desenvolvimento contínuo dos colaboradores. No entanto, a ausência de feedback consistente durante o ciclo avaliativo foi identificada como um ponto crítico, prejudicando o funcionamento do sistema de GD. A falta desse processo contínuo impede correções oportunas e limita as oportunidades de crescimento.

“Então se você tem um chefe que impede seu desenvolvimento e dá nota máxima na avaliação, é mais desmotivador do que receber notas menores de alguém que está claramente interessado no seu crescimento pessoal” (Indivíduo 1).

“As pessoas sentem que os gestores estão mais receptivos, escutando mais e [...] elas recebem um feedback do gestor que antes não recebiam” (Indivíduo 5).

Os relatos demonstram que a adoção de práticas regulares e sinceras de feedback é essencial para que o sistema de GD alcance seus resultados, promovendo tanto a melhoria no desempenho individual quanto o fortalecimento das relações dentro da organização. Esses relatos convergem com a literatura e com os itens propostos para a escala, os quais apontam o apoio e a confiança na relação entre avaliado e avaliador, principalmente o gestor, que exerce um papel central e crucial na efetividade do sistema de GD, conforme apontado por Schleicher et al. (2019).

A categoria **Desempenho no trabalho** abrange comportamentos manifestados no ambiente profissional que visam alcançar resultados organizacionais (Campbell, 1993; Motowilde; Borman; Schmit, 1997). Durante as entrevistas, os participantes destacaram a avaliação de desempenho como uma ferramenta essencial para melhorar o desempenho dos servidores no trabalho.

“Espero que em um ano, no máximo dois, consigamos ter uma avaliação de desempenho voltada para o servidor para que possamos extrair de fato onde o

servidor precisa se aperfeiçoar e onde a organização precisa capacitá-lo para obter um melhor desempenho” (Indivíduo 5).

De forma geral, os relatos apontam que esse critério é central para justificar a adoção do sistema de GD. Os entrevistados enfatizaram a relevância do sistema em promover desempenho superior ao identificar áreas específicas para o desenvolvimento dos servidores. Tal perspectiva reforça o uso da avaliação como uma ferramenta estratégica de gestão de talentos, auxiliando na identificação de onde cada servidor pode contribuir de maneira mais eficaz.

Além disso, os dados oriundos do sistema de GD foram apontados como facilitadores para uma maior eficiência organizacional, possibilitando decisões mais assertivas sobre a realocação de servidores com base em seus perfis e competências.

“Onde é decidido, onde o servidor será realocado com o melhor perfil ou perfil onde ele seria melhor em outra unidade é um processo que oportuniza uma inter-relação grande entre a estrutura de uma organização, se for bem estruturado” (Indivíduo 13).

Essas falas evidenciam que, quando bem implementado, o sistema de GD não apenas impulsiona o desempenho individual, mas também promove maior alinhamento organizacional, contribuindo para uma estrutura mais coesa e eficiente. Portanto, essa categoria ocupa um papel central e relevante na efetividade do sistema de GD e a sua representação na escala de Transferência se restringe à melhoria do desempenho individual do ponto de vista da tarefa.

O critério **Motivação**, relacionado à disposição ou tendência para exercer esforços no trabalho de acordo com os procedimentos e políticas organizacionais (Jiang et al., 2013; Latham et al., 2005), foi mencionado pontualmente, mas com uma reflexão significativa. O relato evidencia a importância da avaliação de desempenho como catalisador de mudanças organizacionais e motivador de crescimento pessoal. Nesse contexto, destaca-se que uma avaliação adequada, mesmo com notas não máximas, pode ser mais motivadora do que uma avaliação complacente, que não reflete o desempenho real.

“A surpresa ao final do ciclo e a falta de feedback ou chance de melhorar antes de tirar uma nota ruim são mais desestimulantes do que uma nota 8,5 ou 9,6 ou 9,8... então se você tem um chefe que impede seu desenvolvimento e dá nota máxima na avaliação, é mais desmotivador do que receber notas menores de alguém que está claramente interessado no seu crescimento pessoal” (Indivíduo 1).

Dessa forma, a efetividade de um sistema de GD é demonstrada pela sua capacidade de identificar necessidades de aprimoramento no desempenho, promovendo uma motivação genuína para o desenvolvimento contínuo no trabalho.

Por fim, percebe-se que os resultados das questões semiestruturadas do roteiro de entrevista forneceram informações que convergem com as principais categorias das dimensões analisadas, mesmo que algumas categorias não tenham sido mencionadas explicitamente pelos entrevistados nos conteúdos analisados. Essa ausência de menção direta pode estar relacionada à natureza mais exploratória das perguntas semiestruturadas, que oferecem maior liberdade aos entrevistados para priorizar os aspectos que consideram mais relevantes.

Antecipando essa possibilidade de omissões, o roteiro foi estrategicamente elaborado para incluir uma parte estruturada (Seção 3), na qual foram apresentados instrumentos contendo as dimensões e os critérios com suas respectivas definições, além de itens específicos para análise e avaliação dos entrevistados. Essa etapa adicional permitiu validar as categorias inicialmente propostas, ampliar a compreensão das percepções dos participantes e assegurar que as dimensões essenciais fossem devidamente consideradas.

Ademais, a estruturação do roteiro contribuiu para complementar as informações obtidas na fase semiestruturada, fornecendo um panorama mais abrangente e detalhado sobre a percepção dos entrevistados em relação à efetividade do sistema de GD. Com essa abordagem metodológica, ao combinar etapas exploratórias e estruturadas, garantiu-se maior robustez à análise e alinhamento com os objetivos da pesquisa, reforçando a adequação dos critérios propostos às práticas e desafios vivenciados nas organizações públicas.

A seguir são apresentados os resultados da análise dos dados estruturados.

4.1.3. Análise de Conteúdo dos Dados Estruturados

É importante destacar que o conteúdo estruturado da Seção 3, além de validar as categorias apontadas anteriormente, também teve como objetivo que os entrevistados avaliassem o conteúdo dos itens traduzidos das escalas revisadas nas literaturas e que originaram a primeira versão dos instrumentos das escalas, submetida à avaliação dos entrevistados. Nesse sentido, na Seção 3 foi solicitado que o entrevistado analisasse item

a item do instrumento apresentado, novamente tomando como referência suas experiências anteriores com a GD na Administração Pública, assim como a sua realidade organizacional. Cada entrevistado pôde discorrer de forma genérica ou específica acerca dos itens da dimensão analisada, indicando a magnitude do atributo a que os itens se referem e sua aplicação no contexto da Administração Pública.

Os itens que não apresentavam dificuldade de compreensão pelos sujeitos eram mantidos no instrumento. Os entrevistados também puderam sugerir alterações no texto do conteúdo dos itens, indicar a necessidade de inclusão de outros aspectos relevantes que não foram contemplados e a exclusão de itens duplicados, confusos ou fora do contexto da Administração Pública. Ao longo das entrevistas, alguns itens foram reformulados com base nas dificuldades de compreensão ou se diferentes daquilo que se pretendia com a escala. Dessa forma, nem todos os itens inicialmente propostos foram submetidos a todos os entrevistados, uma vez que poderiam ter sido excluídos ou sofrido modificações substanciais ao longo das entrevistas realizadas.

Em geral, os entrevistados apontaram uma grande sobreposição e duplicidade entre os itens e necessidades de ajustar a escrita tanto para tornar o conteúdo mais claro quanto para adequar termos à realidade da Administração Pública. Este foi o caso da solicitação de substituição do termo: gerente por chefia, recompensas por gratificações, holerite por contracheque, entre outros. Além desses aprimoramentos, houve também a solicitação de padronização entre os itens, uma vez que grande parte deles foram inicialmente retirados de escalas e instrumentos diferentes. Outro aspecto solicitado e relacionado à padronização refere-se à adequação do tempo verbal dos verbos.

Além da análise posterior a cada entrevista realizada, houve também uma análise global dos itens após a finalização de todas as entrevistas. Assim foi possível identificar os itens que ainda apresentavam duplicidade ou aqueles que não representavam o atributo a que o item se referia. Nessa etapa, praticamente todos os itens também passaram por uma reformulação na sua escrita, tendo em vista a sugestão de alguns entrevistados de que os itens fossem mais padronizados, uma vez que a construção da escala contemplou diferentes instrumentos e, logo, carregavam uma diversidade de estilos de escrita na descrição dos itens.

A Tabela 26 apresenta a quantidade de itens originais da escala, os itens incluídos ou excluídos e o número final de itens em cada um dos critérios das dimensões de avaliação.

Tabela 26

Quantidade itens originais e finais das escalas de Reação, Aprendizagem e Transferência após análise dos entrevistados

Dimensão e Critério	Quantidade original de itens	Quantidade itens incluídos	Quantidade de exclusões	Quantidade total final de itens
<i>Reação</i>	106	5	52	58
Justiça	37	1	7	31
Precisão	15	0	3	12
Afetividade	5	2	0	7
Utilidade	7	2	1	8
Satisfação	42	0	42	0
<i>Aprendizagem</i>	36	0	1	34
Conhecimento	6	1	1	6
Atitude	11	0	1	10
Habilidade	19	0	1	18
<i>Transferência</i>	66	0	0	56
Atitudes no trabalho	31	0	6	25
Motivação	19	0	2	17
Desempenho	8	0	0	8
Relação interpessoal	5	0	0	5
Bem-estar	3	0	2	1

A Escala de Reação, inicialmente composta por 106 itens divididos em cinco critérios, foi estruturada para fornecer uma visão abrangente da variável Reação, amplamente estudada no campo de GD, conforme discutido anteriormente (*e.g.* Awan, 2020; Iqbal et al., 2019; Keeping & Levy, 2000; Kim, 2016; Modipane et al., 2019; Sharma et al., 2016, 2021; Tuytens & Devos, 2012; Wright & Cheung, 2007). Assim, optou-se por iniciar a coleta de dados com esse número de itens justamente para capturar de forma completa os diferentes estudos disponíveis sobre o tema.

A quantidade inicial elevada de itens reflete o alcance e a diversidade de abordagens sobre Reação, além de contemplar o critério de satisfação, o qual possui a maior quantidade de itens desta dimensão. Isso porque a satisfação tem uma característica mais genérica, contemplando diversos aspectos e fazendo com que o critério possa se sobrepor a outros. Por essa razão, essa sobreposição será discutida mais adiante neste estudo.

Durante essa etapa de entrevistas, os participantes sugeriram a inclusão de cinco novos itens à escala de Reação, que não estavam presentes nas escalas descritas na

literatura. Esses itens referem-se aos critérios de justiça, afetividade e utilidade, e abordam aspectos específicos da experiência de GD no setor público brasileiro.

Entre as sugestões, destaca-se a inclusão de um item sobre a percepção de periodicidade adequada para a avaliação, que permitiria tempo suficiente para mudanças comportamentais, sendo essa uma questão ligada ao critério de Justiça. Essa sugestão surgiu principalmente a partir da leitura dos itens de justiça procedimental, relacionados aos procedimentos do sistema de GD (Modipane et al., 2019).

Outro item proposto refere-se à percepção de que as avaliações são frequentemente formais e pouco fidedignas, uma crítica associada ao critério de Utilidade. Esse ponto foi recorrentemente mencionado pelos entrevistados, como apresentado na seção anterior, e representa um aspecto que não havia sido contemplado nos itens previamente traduzidos e adaptados para as escalas. Essas sugestões reforçam a importância de considerar as especificidades do setor público brasileiro na construção de instrumentos mais robustos e aderentes à realidade local.

Por último, em relação à Afetividade, foi abordada a possibilidade de vivenciar tanto emoções positivas quanto receios em relação aos resultados da avaliação. Esse aspecto foi mencionado pelos entrevistados durante a leitura do instrumento na Seção 3. Observou-se que havia uma predominância de itens associados a aspectos negativos e destacaram que seria possível sentir-se feliz com os resultados obtidos, além de ser normal experimentar receio com relação aos resultados, ambos dependendo do contexto e da interpretação dos mesmos.

Ao final do processo, a escala de Reação foi reduzida para 58 itens, com a exclusão de 52 itens (49% de itens excluídos). Essa redução significativa deve-se, em grande parte, à opção de desconsiderar o critério de Satisfação na variável Reação. Durante as entrevistas, ficou evidente o desconforto dos participantes em relação à sobreposição dos itens de satisfação, que muitas vezes mediam aspectos já abordados em outros critérios de Reação. Isso ocorre porque o critério de satisfação tem uma natureza tipicamente genérica, o que resulta em operacionalizações inconsistentes e influenciadas por outras variáveis (Keeping & Levy, 2000). Embora, inicialmente, tenha sido feita a escolha teórica de manter esse critério no modelo, optou-se por excluí-lo, juntamente de seus 42 itens, para garantir a exclusividade dos critérios restantes e torná-los mais objetivos e consistentes.

A escala de satisfação continua sendo relevante nos estudos de avaliação de efetividade da GD, especialmente como um critério inicial de análise. Contudo, com a

evolução dos modelos de avaliação, especialmente os modelos integrativos, a forma isolada de mensuração da satisfação pode se tornar inconsistente e potencialmente confusa. Por isso, a escala de satisfação não deve ser desconsiderada, mas recomenda-se que seja analisada de forma independente dos demais critérios ou evitando sobreposições que possam comprometer a objetividade da avaliação quando combinada com outros critérios.

Além da exclusão dos itens de satisfação, itens dos demais critérios foram excluídos por duplicidade de informações, especialmente no critério de justiça. Observou-se que as escalas de justiça variavam quanto ao nível de detalhamento, com algumas abordando justiça em seus diferentes tipos e outras de forma mais agregada. De modo geral, os entrevistados preferiram avaliar a justiça de forma global, sem detalhar aspectos específicos do processo de GD.

Diferentemente da escala de Reação, as escalas de Aprendizagem e Transferência, por serem baseadas em estudos menos consolidados, foram submetidas com um número menor de itens. A escala de Aprendizagem, em particular, conta com três critérios – Conhecimento, Habilidade e Atitude – e iniciou com 36 itens. Essa escala é a menos consolidada e aborda uma dimensão relativamente nova no campo (Schleicher et al., 2019). Apesar disso, os itens descritos foram bem recebidos pelos entrevistados e a escala passou por poucas modificações, resultando na exclusão de apenas três itens, um em cada critério. No critério de Conhecimento, com base na definição dessa categoria, alguns entrevistados apontaram a ausência de um item mais específico e explícito que abordasse a percepção de adquirir as habilidades necessárias para contribuir de forma efetiva e agregar valor à organização. Essa lacuna foi considerada relevante, resultando na inclusão desse aspecto no questionário. Com esse ajuste, a escala de Aprendizagem foi consolidada, totalizando 34 itens.

A escala de Transferência foi inicialmente composta por 66 itens, distribuídos em cinco critérios: Atitudes no Trabalho, Bem-estar, Desempenho, Motivação e Relação Interpessoal. O critério com a maior quantidade de itens foi Atitudes no Trabalho, pois abrange temas como comprometimento organizacional, engajamento e satisfação no trabalho, áreas com uma base teórica e empírica mais consolidada. Por essa razão, apenas seis itens desse critério foram excluídos, sem necessidade de inclusão de novos itens. Em contraste, o critério de Bem-estar, que contava com apenas três itens e apresentou dificuldades em encontrar itens específicos na literatura, foi reduzido a um único item, já que os outros dois foram considerados redundantes pelos entrevistados. Por fim, dois itens

de Motivação foram excluídos. Os demais critérios permaneceram inalterados. Assim, a escala de Transferência foi finalizada com 56 itens.

Portanto, essa etapa de submissão dos itens traduzidos, originalmente desenvolvidos para outros contextos, aos praticantes permitiu avaliar sua aplicabilidade à realidade da Administração Pública brasileira. Esse processo preliminar foi essencial para identificar elementos específicos que poderiam não se ajustar ao contexto público nacional, como terminologias específicas. Realizar essa etapa antes da avaliação pelos juízes acadêmicos evitou o esforço desnecessário de análise de itens potencialmente inadequados, garantindo uma maior aderência ao ambiente de trabalho dos servidores públicos. Enquanto os juízes acadêmicos aplicam critérios científicos com pretensões mais universais, validar inicialmente os itens com o público praticante trouxe uma perspectiva prática indispensável, conferindo maior relevância e a precisão da escala final.

Com as escalas ajustadas pelos participantes do campo, procedeu-se com a validação de juízes.

4.1.4. Validação de juízes

A validação de juízes ocorreu para cada uma das três escalas com intuito de refinar o instrumento de pesquisa resultante da fase anterior. Essa validação se deu partir da análise de conteúdo, relativa à adequação do item ao fator, representada pela Taxa de Concordância de Juízes (TCJ), e da análise semântica do item, referente à clareza e chamado de Coeficiente de Validade de Conteúdo (CVC). Neste estudo, o índice CVC se refere exclusivamente à clareza da linguagem dos itens. Em outros estudos, ele pode abranger também a pertinência prática e a relevância teórica (Cassepp-Borges et al., 2010). A pertinência prática, por sua vez, pode ser verificada por meio de análise fatorial exploratória, a qual foi realizada posteriormente. Já a relevância teórica se distingue do TCJ, pois mede a concordância entre os juízes em uma escala (Borsa & Seize, 2017; Cassepp-Borges et al., 2010), o que, nesse caso, não foi considerado necessário para esta pesquisa.

O resultado obtido a partir da validação de juízes pode ser conferido na Tabela 27, a qual reúne a quantidade de itens originalmente submetidos à análise e a quantidade itens finais, após a exclusão dos itens que não obtiveram adequação ao seu fator e/ou por ausência de clareza.

Tabela 27

Quantidade itens originais e finais das escalas de Reação, Aprendizagem e Transferência após análise dos juízes.

Escala	Quantidade original de itens	Quantidade de itens excluídos	Quantidade final de itens
Reação	58	39	19
Aprendizagem	34	11	23
Transferência	56	31	25

As escalas de Reação e Transferência apresentaram uma maior quantidade de itens não validados em comparação à escala de Aprendizagem. Isso se deve, em parte, ao fato de essas escalas conterem um número maior de itens, reflexo de seus critérios serem mais amplamente estudados e consolidados na literatura do que os itens de Aprendizagem.

É importante pontuar que a exclusão de uma quantidade significativa de itens na etapa de validação pelos juízes, mesmo após a validação inicial com o público-alvo, mostra-se aceitável. Isso ocorre porque a construção das escalas envolveu a adaptação de itens traduzidos de diferentes instrumentos, conforme detalhado nas etapas anteriores. Essa abordagem foi adotada intencionalmente para abranger a maior variedade possível de conteúdos associados às diferentes teorias alternativas conhecidas sobre os critérios de cada escala. Esse processo de incluir itens variados visou criar um instrumento inicial mais amplo e abrangente do que a visão teórica restrita oferece.

Contudo, o maior volume de itens também contribui para uma maior ocorrência de conteúdos sobrepostos entre os itens das escalas, especialmente considerando que a literatura sobre a efetividade da GD apresenta uma multiplicidade de conceitos e medidas frequentemente confusas, o que dificulta a mensuração clara dessas dimensões. Essa complexidade foi ressaltada pelos juízes, que apontaram dificuldades em validar itens redundantes ou com interpretações ambíguas. Como parte do processo de refinamento, esses itens foram eliminados durante a validação, assegurando que o instrumento final refletisse os critérios com maior clareza, objetividade e consistência (Clark & Watson, 1995).

As subseções a seguir detalham as razões das alterações e apresentam os resultados obtidos na validação realizada pelos juízes para cada uma das escalas.

Escala de Percepção de Reação

A escala de Reação foi inicialmente composta por 106 itens traduzidos da literatura revisada, distribuídos em cinco critérios: Afetividade, Justiça, Precisão, Satisfação e Utilidade. Após a submissão desses itens à população-alvo e a incorporação das contribuições do campo, a escala foi reduzida a 58 itens, então organizada em quatro critérios, excluindo a Satisfação, conforme razões discutidas anteriormente. Portanto, essa escala foi submetida à validação dos juízes com quatro critérios e teve 19 itens validados, conforme razões apresentadas na Tabela 28.

Tabela 28

Validação de juízes da Escala de Percepção de Reação

n	Itens	Dimensão	TCJ	CVC	Status do item - Justificativa de exclusão
1	Sinto que a avaliação que recebo do meu desempenho é condizente com o que eu mereço	Justiça	67%	0,63	Excluído - CVC < 0,7
2	Acho a perturbador ter meu desempenho avaliado	Afetividade	100%	0,63	Excluído - CVC < 0,7
3	Acredito que a avaliação de desempenho me ajuda a aprender como posso fazer melhor o meu trabalho	Utilidade	100%	0,63	Excluído - CVC < 0,7
4	Acredito que a avaliação de desempenho me auxilia a enxergar chances de crescimento na organização	Utilidade	100%	0,70	Validado
5	Acredito que a avaliação de desempenho que recebi corresponde a minha autoavaliação	Precisão	100%	0,63	Excluído - CVC < 0,7
6	Acredito que a avaliação de desempenho que recebo considera objetivamente o meu desempenho atual em relação ao esperado	Precisão	67%	0,70	Excluído - CVC < 0,7
7	Acredito que a avaliação de desempenho realizada representa uma aferição precisa do meu desempenho	Precisão	100%	0,56	Excluído - CVC < 0,7
8	Acredito que a avaliação de desempenho respeita padrões éticos e morais	Justiça	100%	0,76	Validado
9	Acredito que a avaliação do meu desempenho me ajuda a entender meus erros	Utilidade	100%	0,63	Excluído - CVC < 0,7

10	Acredito que a avaliação que recebo do meu desempenho é justa	Justiça	100%	0,76	Validado
11	Acredito que a avaliação que recebo do meu desempenho fornece uma avaliação adequada do trabalho que realizo	Justiça	100%	0,76	Validado
12	Acredito que a avaliação que recebo do meu desempenho me fornece uma boa avaliação do esforço que coloco no meu trabalho	Justiça	100%	0,76	Validado
13	Acredito que a minha chefia imediata avalia meu desempenho de forma educada	Justiça	67%	0,70	Excluído - CVC < 0,7
14	Acredito que as minhas metas ou objetivos são precisamente avaliados como parte do processo de avaliação do desempenho	Precisão	100%	0,50	Excluído - CVC < 0,7
15	Acredito que eu aprendo muito com a avaliação de desempenho que recebo	Utilidade	100%	0,83	Validado
16	Acredito que minha chefia imediata avalia meu desempenho de forma justa	Justiça	67%	0,70	Excluído - CVC < 0,7
17	Acredito que recebo feedback construtivo sobre o meu desempenho no trabalho por parte do(s) meu(s) avaliador(es)	Justiça	0%	0,76	Excluído - TCJ < 0,6%
18	Acredito que recebo feedback oportuno sobre o meu desempenho no trabalho por parte da minha chefia imediata	Justiça	33%	0,70	Excluído - TCJ < 0,6%
19	Acredito que as minhas recompensas estão relacionadas a avaliação do desempenho que recebi	Precisão	0%	0,50	Excluído - TCJ < 0,6% e clareza < 0,7
20	Compreendo as razões da avaliação de desempenho que recebo	Justiça	67%	0,56	Excluído - CVC < 0,7
21	Compreendo o que devo fazer para receber uma avaliação de desempenho melhor	Justiça	33%	0,70	Excluído - TCJ < 0,6%
22	Compreendo que o feedback contínuo me fornece uma avaliação precisa de como estou me saindo em relação ao desempenho esperado	Precisão	67%	0,50	Excluído - CVC < 0,7
23	Concentro melhor meus esforços a partir da identificação das metas relevantes para o	Precisão	67%	0,63	Excluído - CVC < 0,7

	atingimento dos objetivos da organização				
24	Confio que a minha chefia imediata comunica claramente o desempenho esperado	Justiça	33%	0,63	Excluído - TCJ < 0,6% e clareza < 0,7
25	Conheço as motivações utilizadas pela minha chefia imediata para avaliar meu desempenho	Justiça	33%	0,63	Excluído - TCJ < 0,6% e clareza < 0,7
26	Considero importante minha avaliação de desempenho ser atrelada a remuneração que recebo	Utilidade	67%	0,56	Excluído - CVC < 0,7
27	Considero que a avaliação de desempenho que recebo é livre de vieses	Justiça	100%	0,56	Excluído - CVC < 0,7
28	Considero que a avaliação de desempenho que recebo é precisa em relação ao meu desempenho	Precisão	100%	0,90	Validado
29	Considero que a avaliação do meu desempenho reflete o meu verdadeiro desempenho	Precisão	100%	0,70	Validado
30	Considero que a avaliação que recebo do meu desempenho afere o que tenho contribuído para a organização	Justiça	0%	0,70	Excluído - TCJ < 0,6%
31	Considero que a minha chefia imediata avalia meu desempenho com dignidade	Justiça	67%	0,56	Excluído - CVC < 0,7
32	Considero que a minha chefia imediata avalia o desempenho dos funcionários de forma justa	Justiça	100%	0,76	Validado
33	Considero que a minha chefia imediata se abstém de observações ou comentários impróprios ao avaliar meu desempenho	Justiça	67%	0,50	Excluído - CVC < 0,7
34	Considero que a periodicidade com que a avaliação de desempenho é aplicada é apropriada para capturar o meu desempenho de forma fidedigna	Justiça	33%	0,83	Excluído - TCJ < 0,6%
35	Considero que o feedback anual que recebo é uma representação precisa dos feedbacks contínuos recebidos ao longo daquele ciclo de desempenho	Precisão	67%	0,63	Excluído - CVC < 0,7
36	Considero que o processo de avaliação do desempenho me ajuda a me desenvolver como profissional	Utilidade	100%	0,90	Validado

37	Consigo visualizar como a minha avaliação de desempenho se compara a das outras pessoas em minha organização que realizam atividades semelhantes	Justiça	33%	0,43	Excluído - TCJ < 0,6% e CVC < 0,7
38	Estou de acordo com a avaliação final que recebo do meu desempenho	Justiça	67%	0,63	Excluído - CVC < 0,7
39	Fico à vontade ao ter meu desempenho avaliado	Afetividade	100%	0,63	Excluído - CVC < 0,7
40	Fico feliz em receber o resultado da minha avaliação de desempenho	Afetividade	100%	0,96	Validado
41	Fico tenso(a) ao ter meu desempenho avaliado*	Afetividade	100%	0,70	Validado
42	Identifico que são utilizadas medidas objetivas para avaliar meu desempenho	Justiça	0%	0,56	Excluído - TCJ < 0,6% e CVC < 0,7
43	Minha chefia imediata e eu atualizamos minhas metas conforme os objetivos da organização mudam	Precisão	100%	0,70	Validado
44	Participo do estabelecimento dos indicadores que são utilizados para avaliar meu desempenho no trabalho	Justiça	67%	0,96	Validado
45	Percebo que a avaliação de desempenho me fez refletir sobre o meu desempenho	Utilidade	100%	0,83	Validado
46	Percebo que as recompensas que recebo são baseadas no meu desempenho	Justiça	100%	0,83	Validado
47	Posso recorrer da avaliação de desempenho que recebo caso eu não concorde	Justiça	100%	0,50	Excluído - CVC < 0,7
48	Recebo orientação ou assistência para me ajudar a melhorar meu desempenho por parte da minha chefia imediata	Justiça	33%	0,76	Excluído - TCJ < 0,6%
49	Sinto que a avaliação de desempenho me causa estresse*	Afetividade	100%	0,76	Validado
50	Sinto que a avaliação de desempenho na organização é utilizada para além de objetivos administrativos e proforma	Utilidade	100%	0,56	Excluído - CVC < 0,7
51	Sinto que a avaliação que recebo é justificada dado o meu desempenho	Justiça	0%	0,63	Excluído - TCJ < 0,6% e CVC < 0,7
52	Sinto que durante o período de avaliação, minhas possibilidades	Precisão	0%	0,56	Excluído - TCJ < 0,6% e CVC < 0,7

	de melhoria são claramente apontadas para mim					
53	Sinto que exerço influência no processo de avaliação de desempenho da minha organização	Justiça	33%	0,50	Excluído - TCJ < 0,6% e CVC < 0,7	
54	Sinto raiva ao ter meu desempenho avaliado*	Afetividade	100%	0,76	Validado	
55	Sinto-me confortável em expressar minhas opiniões e sentimentos sobre o processo de avaliação de desempenho da minha organização	Justiça	33%	0,83	Excluído - TCJ < 0,6%	
56	Sinto-me informado pela minha chefia imediata sobre o quanto bem estou desempenhando minhas atividades	Justiça	100%	0,76	Validado	
57	Tenho oportunidades suficientes (como atividades ou projetos desafiadores) para obter uma avaliação de desempenho melhor	Justiça	67%	0,63	Excluído - CVC < 0,7	
58	Tenho receio sobre o que a avaliação de desempenho pode indicar*	Afetividade	100%	0,63	Excluído - CVC < 0,7	

Nota: índices inadequados aparecem em negrito. Alguns itens com 0,70 de CVC foram excluídos, pois apresentavam valor menor que 0,7 quando consideradas as casas decimais.

Os juízes concordaram sobre a adequação ao fator em 43 itens, mas divergiram em 15, com as discordâncias concentradas principalmente nas subdimensões de Justiça e Precisão. Esse contraste pode refletir uma prática identificada em estudos recentes (como Sharma et al., 2016), que tendem a tratar Justiça e Precisão como critérios independentes de efetividade, apesar de muitas vezes abordarem aspectos conceitualmente interligados. Na seção anterior, vários entrevistados indicaram que alguns itens de Justiça pareciam avaliar Precisão e vice-versa, o que ressalta uma ambiguidade na interpretação desses construtos, a qual ficou aparente pela avaliação fornecida pelos juízes.

Dos 43 itens considerados adequados ao fator, 24 não alcançaram o índice mínimo de clareza, de acordo com o CVC. Devido à ausência de comentários específicos dos juízes para orientar revisões nesses itens, eles foram excluídos da escala para não prejudicar a validade da escala. A dificuldade de clareza pode estar associada à sobreposição conceitual observada na literatura. Como não há um modelo integrativo, os estudos utilizam escalas distintas, mas que frequentemente avaliam temáticas semelhantes, tornando alguns itens cansativos ou vagos e carecendo de uma distinção

clara dos demais. Em outras palavras, apesar de adequados ao fator, esses itens provavelmente apresentavam redundância de informações, conforme indicado anteriormente pelos entrevistados. Tal redundância dificulta uma avaliação precisa e poderia levar a respostas inadequadas à escala.

Ademais, um número reduzido de juízes pode amplificar o efeito de notas baixas atribuídas por um único juiz (*outlier*), impactando significativamente o índice CVC de cada item (Carmines & Zeller, 1979). Embora fosse viável ajustar esses itens para melhorar sua clareza e submetê-los novamente à avaliação, optou-se por uma abordagem parcimoniosa, privilegiando uma escala mais enxuta, com itens de maior clareza e sem sobreposições de conteúdo. Assim, ao todo foram excluídos 39 itens.

Portanto, a validação dos juízes da escala de percepção de Reação resultou em um instrumento composto por 19 itens, distribuídos da seguinte forma: 4 itens para a dimensão “Afetividade”, 8 itens para a dimensão “Justiça”, 3 itens para a dimensão “Precisão” e 4 itens para a dimensão “Utilidade”. Todos os quatro critérios previstos inicialmente foram mantidos e representados pelo rol de itens. Dessa forma, considerou-se que o questionário de Reação continuou contemplando e representando as respectivas definições constitutivas, não havendo prejuízo conceitual ou de construto que justificasse uma nova análise teórica ou uma nova validação de juízes.

Escala de Percepção de Aprendizagem

A Escala de Percepção de Aprendizagem foi composta por três critérios: Atitude, Conhecimento e Habilidade, e inicialmente contemplava 36 itens. As análises da aplicabilidade com o público-alvo mantiveram a escala com os três critérios: Atitude, Conhecimento e Habilidade, mas agora com 34 itens, os quais foram submetidos à análise dos juízes, ficando com 23 itens validados, conforme a Tabela 29.

Tabela 29

Validação de juízes da Escala de Percepção de Aprendizagem

n	Itens	Dimensão	TCJ	CVC	Status do item - Justificativa de exclusão
1	Considero que a avaliação de desempenho contribui para o meu desenvolvimento profissional	Habilidade	33%	0,96	Excluído - TCJ < 0,6%

2	Acho que a avaliação de desempenho é interessante	Atitude	100%	0,63	Excluído - CVC < 0,7
3	Acho que o sistema de GD/AD é bom para mim	Atitude	100%	0,63	Excluído - CVC < 0,7
4	Acredito que a avaliação de desempenho esclarece as expectativas em relação ao meu trabalho	Conhecimento	67%	0,96	Validado
5	Acredito que a avaliação de desempenho fornece informações sobre minhas contribuições pessoais e meu valor na organização	Conhecimento	67%	0,83	Validado
6	Adquiro o conhecimento necessário para executar minhas tarefas relacionadas ao meu papel na avaliação de desempenho	Habilidade	33%	0,90	Excluído - TCJ < 0,6%
7	Considero que a avaliação de desempenho é importante para mim	Atitude	50%	0,70	Excluído - TCJ < 0,6%
8	Considero que o processo de avaliação de desempenho é agradável	Atitude	100%	0,90	Validado
9	Considero que participar da avaliação de desempenho é algo que eu tenho que fazer	Atitude	67%	0,83	Validado
10	Considero que participar da avaliação de desempenho é divertido	Atitude	100%	0,96	Validado
11	Considero-me competente para executar minhas tarefas relacionadas ao meu papel na avaliação de desempenho	Habilidade	67%	0,90	Validado
12	Consigo receber ou fornecer informações claras e diretas sobre o desempenho esperado	Habilidade	100%	0,90	Validado
13	Domino melhor as tarefas relacionadas ao meu papel na avaliação de desempenho	Habilidade	33%	0,83	Excluído - TCJ < 0,6%
14	Entendo as razões pelas quais eu devo participar da avaliação de desempenho	Atitude	0%	0,96	Excluído - TCJ < 0,6%
15	Entendo que pode haver boas razões para contribuir com a avaliação de desempenho, mas pessoalmente não vejo nenhuma*	Atitude	0%	0,50	Excluído - TCJ < 0,6% e CVC < 0,7
16	Minhas tarefas ficam claras para mim	Conhecimento	100%	0,76	Validado
17	Obtenho as habilidades necessárias para executar as tarefas relacionadas ao meu papel na avaliação de desempenho	Conhecimento	100%	0,90	Validado

18	Participo da avaliação de desempenho para o meu próprio bem	Atitude	100%	0,90	Validado
19	Sei como receber ou fornecer feedback mesmo se não forem momentos formais de feedback ou de avaliação do desempenho	Habilidade	67%	0,90	Validado
20	Sei o que se espera de mim em relação à execução das tarefas	Conhecimento	67%	0,83	Validado
21	Sei quais são minhas responsabilidades	Conhecimento	100%	0,63	Excluído - CVC < 0,7
22	Sinto que tenho que participar da avaliação de desempenho	Atitude	100%	0,83	Validado
23	Sinto-me bem ao participar da avaliação de desempenho	Atitude	100%	0,96	Validado
24	Sinto-me capaz de fazer mudanças em meu desempenho geral do trabalho baseadas na minha avaliação	Habilidade	100%	0,83	Validado
25	Sinto-me capaz de receber ou fornecer feedback frequente e contínuo sobre o trabalho realizado	Habilidade	100%	0,90	Validado
26	Sinto-me capaz de reconhecer o que fiz bem e o que poderia ter feito melhor no trabalho	Habilidade	33%	0,90	Excluído - TCJ < 0,6%
27	Sinto-me disposto(a) a participar da avaliação de desempenho por decisão pessoal	Atitude	67%	0,76	Validado
28	Sinto-me disposto(a) a participar da avaliação de desempenho porque eu não tenho escolha*	Atitude	67%	0,70	Validado
29	Sinto-me disposto(a) a participar da avaliação de desempenho porque sou obrigado(a)*	Atitude	100%	0,76	Validado
30	Sinto-me disposto(a) a participar da avaliação de desempenho, mas não tenho certeza se é bom realizá-la*	Atitude	100%	0,70	Validado
31	Sinto-me disposto(a) a participar da avaliação de desempenho, mas não tenho certeza se vale a pena*	Atitude	100%	0,70	Validado
32	Sinto-me treinado para executar minhas atribuições relacionadas ao meu papel na avaliação de desempenho	Habilidade	67%	0,76	Validado
33	Tenho a intenção de melhorar meu desempenho baseado na avaliação que recebi	Atitude	33%	0,90	Excluído - TCJ < 0,6%
34	Vejo os benefícios que a avaliação de desempenho me traz	Atitude	67%	0,90	Validado

Dos 34 itens avaliados, os juízes concordaram sobre a adequação ao fator em 26 deles e discordaram em oito. Entre os itens adequados, três não apresentaram clareza suficiente na opinião dos juízes e foram retirados da escala. Ao todo, 11 itens foram excluídos, restando 23 itens validados e relacionados aos três critérios: Atitude, Conhecimento e Habilidade.

Um dos juízes optou por não indicar a adequação dos itens de números 2, 3, 7, 15 e 21 a qualquer fator, justificando que nenhum deles era aplicável aos itens avaliados. Dessa forma, esses itens foram analisados apenas por dois juízes. Desses, dois itens foram validados com 100% de concordância entre os dois avaliadores, enquanto nos outros três itens não houve consenso.

Portanto, o instrumento validado pelos juízes ficou com 23 itens, distribuídos da seguinte forma: 12 itens para a dimensão “Atitude”, 5 itens para a dimensão “Conhecimento”, e 6 itens para a dimensão “Habilidade”. Essas modificações nos itens não afetaram sua interpretação original, tornando desnecessária a realização de uma nova análise teórica ou uma nova validação por parte dos juízes.

Escala de Percepção de Transferência

A Escala de Percepção de Transferência foi inicialmente proposta de forma teórica com 66 itens. Para a validação de juízes, o instrumento contou com 56 itens divididos em cinco critérios: Atitude no Trabalho, Bem-estar, Desempenho no Trabalho, Motivação e Relação Interpessoal. Após essa etapa, a escala contou com 25 itens validados, conforme disposto na Tabela 30.

Tabela 30

Validação de juízes da Escala de Percepção de Transferência

n	Itens	Dimensão	TCJ	CVC	Status do item - Justificativa de exclusão
1	Acredito que o tempo voa quando estou trabalhando	Atitude no trabalho	67%	0,96	Validado
2	Considero a qualidade do meu trabalho excelente	Desempenho no trabalho	100%	0,96	Validado
3	Considero meu trabalho uma motivação em si	Motivação	100%	0,90	Validado

4	Considero que a minha equipe é feita por relacionamentos genuínos	Relações interpessoais	100%	0,63	Excluído - CVC < 0,7
5	Considero que as tarefas que realizo no trabalho representam, elas próprias, uma força motriz no meu trabalho	Motivação	67%	0,63	Excluído - CVC < 0,7
6	Considero que as tarefas que realizo no trabalho são prazerosas	Motivação	33%	0,96	Excluído - TCJ < 0,66%
7	Considero que este trabalho preenche meus planos de carreira	Motivação	100%	0,96	Validado
8	Considero que este trabalho se encaixa nos meus valores pessoais	Motivação	0%	0,76	Excluído - TCJ < 0,66%
9	Considero que o meu trabalho é inspirador	Atitude no trabalho	33%	0,96	Excluído - TCJ < 0,66%
10	Considero que o meu trabalho é útil	Atitude no trabalho	67%	0,96	Validado
11	Considero que o meu trabalho me motiva	Motivação	100%	0,96	Validado
12	Considero que o meu trabalho me permite ganhar muito dinheiro	Motivação	33%	0,96	Excluído - TCJ < 0,66%
13	Considero que o trabalho é como um hobby para mim	Motivação	33%	0,96	Excluído - TCJ < 0,66%
14	Considero que os problemas desta organização são meus	Atitude no trabalho	100%	0,63	Excluído - CVC < 0,7
15	Desempenho melhor minhas funções	Desempenho no trabalho	100%	0,63	Excluído - CVC < 0,7
16	Divirto-me fazendo meu trabalho	Motivação	33%	0,96	Excluído - TCJ < 0,66%
17	Empenho-me na realização do meu trabalho	Desempenho no trabalho	100%	0,96	Validado
18	Esqueço tudo o que se passa a minha volta quando estou trabalhando	Atitude no trabalho	100%	0,76	Validado
19	Estou bastante satisfeito(a) com o meu trabalho	Atitude no trabalho	100%	0,96	Validado
20	Estou entusiasmado(a) com o meu trabalho	Atitude no trabalho	33%	0,96	Excluído - TCJ < 0,66%
21	Estou imerso(a) no meu trabalho	Atitude no trabalho	100%	0,96	Validado
22	Estou interessado(a) no meu emprego atual	Atitude no trabalho	67%	0,83	Validado
23	Estou orgulhoso(a) do que faço neste trabalho	Atitude no trabalho	67%	0,96	Validado
24	Faço este trabalho pelo contracheque	Motivação	100%	0,63	Excluído - CVC < 0,7
25	Faço um esforço extra na realização do meu trabalho	Desempenho no trabalho	100%	0,96	Validado
26	Gosto de conversar sobre a minha organização com pessoas de fora dela	Atitude no trabalho	67%	0,96	Validado

27	Gosto de trabalhar aqui	Atitude no trabalho	100%	0,96	Validado
28	Gosto muito do meu trabalho	Atitude no trabalho	100%	0,96	Validado
29	Identifico efeitos claros no meu desempenho	Desempenho no trabalho	100%	0,56	Excluído - CVC < 0,7
30	Posso receber ordens e pedidos mais claros mesmo que sejam de duas ou mais pessoas	Relações interpessoais	67%	0,50	Excluído - CVC < 0,7
31	Sinto dificuldade em me desligar do trabalho	Atitude no trabalho	33%	0,96	Excluído - TCJ < 0,66%
32	Sinto que "deixo-me ir" quando estou trabalhando	Atitude no trabalho	67%	0,96	Validado
33	Sinto que a minha reputação depende do meu trabalho	Motivação	0%	0,90	Excluído - TCJ < 0,66%
34	Sinto que a organização representa um grande significado pessoal para mim	Atitude no trabalho	67%	0,90	Validado
35	Sinto que este trabalho me traz momentos de prazer	Motivação	0%	0,90	Excluído - TCJ < 0,66%
36	Sinto que meu crescimento e desenvolvimento profissional são apoiados	Relações interpessoais	0%	0,96	Excluído - TCJ < 0,66%
37	Sinto que meu trabalho vale a pena	Atitude no trabalho	67%	0,83	Validado
38	Sinto que o meu trabalho é a minha vida	Motivação	33%	0,83	Excluído - TCJ < 0,66%
39	Sinto que o meu trabalho é desafiante para mim	Atitude no trabalho	0%	0,90	Excluído - TCJ < 0,66%
40	Sinto que o meu trabalho me proporciona um certo padrão de vida	Motivação	33%	0,70	Excluído - TCJ < 0,66%
41	Sinto que tenho que "dar" o melhor no meu trabalho	Motivação	0%	0,90	Excluído - TCJ < 0,66%
42	Sinto que trabalhar é instigante	Motivação	100%	0,63	Excluído - CVC < 0,7
43	Sinto sensações positivas a partir do meu trabalho	Bem-estar	100%	0,83	Validado
44	Sinto um forte sentimento de pertencimento à minha organização	Atitude no trabalho	33%	0,90	Excluído - TCJ < 0,66%
45	Sinto-me "parte da família" na minha organização	Atitude no trabalho	0%	0,90	Excluído - TCJ < 0,66%
46	Sinto-me "emocionalmente ligado(a)" a esta organização	Atitude no trabalho	33%	0,90	Excluído - TCJ < 0,66%
47	Sinto-me capaz de trabalhar por períodos muito longos	Atitude no trabalho	33%	0,90	Excluído - TCJ < 0,66%
48	Sinto-me cheio(a) de energia quando estou trabalhando	Atitude no trabalho	33%	0,96	Excluído - TCJ < 0,66%
49	Sinto-me feliz quando trabalho intensamente	Atitude no trabalho	0%	0,70	Excluído - TCJ < 0,66%

50	Sinto-me sortido(a) por ser pago por um trabalho que eu gosto tanto	Motivação	33%	0,90	Excluído - TCJ < 0,66%
51	Sou auxiliado(a) em relação a obter os recursos de que preciso para fazer meu trabalho	Relações interpessoais	67%	0,83	Validado
52	Sou perseverante no trabalho, mesmo quando as coisas não estão a correr bem	Atitude no trabalho	67%	0,63	Excluído - CVC < 0,7
53	Tenho clareza sobre o quão aceitável meu trabalho é para o meu chefe	Relações interpessoais	67%	0,76	Validado
54	Tenho um desempenho melhor do que o esperado	Desempenho no trabalho	67%	0,90	Validado
55	Tenho um desempenho melhor do que o que pode ser caracterizado como desempenho aceitável	Desempenho no trabalho	67%	0,76	Validado
56	Tento trabalhar o máximo possível	Desempenho no trabalho	67%	0,83	Validado

Uma parcela significativa dos itens foi eliminada devido à discordância entre os juízes quanto à sua classificação nas respectivas dimensões. Foram retirados 22 itens relacionados aos critérios de Motivação e Atitude no Trabalho e apenas um item referente a Relações Interpessoais foi excluído. Ressalta-se que tanto Motivação quanto Atitude no Trabalho eram critérios que possuíam maior número de escalas consolidadas na literatura, o que permitiu incluir um volume mais significativo de itens no questionário inicial.

No entanto, a construção do critério Atitude no Trabalho neste estudo, ao abranger aspectos como engajamento, comprometimento organizacional e satisfação no trabalho, pode ter contribuído para a confusão conceitual observada pelos juízes. Essa integração de múltiplos construtos pode resultar em uma estrutura potencialmente ambígua, dificultando a atribuição clara de determinados itens à dimensão correspondente (Pasquali, 2010). Tal ambiguidade pode ter impacto negativo na avaliação dos juízes sobre a adequação dos itens.

Com relação ao critério Motivação, observou-se que a maioria dos estudos na literatura tende a abordar o tema de maneira genérica, avaliando a motivação no contexto do sistema de gestão de desempenho sem especificar claramente sua direção ou nuances (*e.g.*, Ishaq et al., 2009, 2013; Longenecker et al., 1988; Longenecker & Nykodym, 1996; Maley; Moeller, 2012; Monsur; Akkas, 2015; Roberts, 1992, 1995; Schleicher et al., 2019; Yang & Hsieh, 2007; Yu et al., 2018). Essa falta de especificidade pode ter dificultado a interpretação dos itens com clareza suficiente para evitar interpretações múltiplas ou contraditórias. Portanto, a exclusão de itens relacionados a Motivação e Atitude no Trabalho favoreceu um maior rigor na definição conceitual e no delineamento

dos itens da escala, especialmente para esses casos de múltiplos aspectos integrados sob uma mesma dimensão.

Por fim, a validação de juízes acarretou a exclusão de 31 itens, resultando em um instrumento composto por 25 itens, distribuídos da seguinte forma: 13 itens para a dimensão “Atitude no trabalho”, 1 item para “Bem-estar”, 6 itens para a dimensão “Desempenho no trabalho”, e 3 itens para a dimensão “Motivação”, e 2 itens para “Relações interpessoais”. Essas modificações nos itens não afetaram sua interpretação original, tornando desnecessária a realização de uma nova análise teórica ou de construto.

Ao final da validação de todas as escalas, foram analisadas as observações e sugestões de alterações semânticas apresentadas pelos juízes. Embora poucas sugestões tenham sido feitas, elas contribuíram para aprimorar a clareza dos itens validados. Adicionalmente, a pesquisadora realizou uma revisão detalhada dos itens com o objetivo de padronizar e uniformizar as escalas, assegurando maior consistência nos itens. Os itens após os ajustes semânticos estão descritos na Tabela 31.

Tabela 31

Itens das escalas de Reação, Aprendizagem e Transferência após ajustes semânticos

n	Variável	Item avaliado pelo juiz	Item após incorporar ajustes semânticos
<i>Escala de Percepção de Reação</i>			
1	AFET1	Fico feliz em receber o resultado da minha avaliação de desempenho	Fico feliz em receber o resultado da minha avaliação de desempenho
2	AFET2*	Fico tenso(a) ao ter meu desempenho avaliado*	Fico tenso(a) ao ter meu desempenho avaliado
3	AFET3*	Sinto que a avaliação de desempenho me causa estresse*	A avaliação de desempenho me causa estresse
4	AFET4*	Sinto raiva ao ter meu desempenho avaliado*	Sinto raiva ao ter meu desempenho avaliado
5	JUST1	Acredito que a avaliação de desempenho respeita padrões éticos e morais	A avaliação de desempenho que recebo respeita os padrões éticos e morais
6	JUST2	Acredito que a avaliação que recebo do meu desempenho é justa	A avaliação que recebo do meu desempenho é equânime
7	JUST3	Acredito que a avaliação que recebo do meu desempenho fornece uma avaliação adequada do trabalho que realizo	A avaliação que recebo do meu desempenho me fornece uma avaliação adequada do trabalho que eu realizo
8	JUST4	Acredito que a avaliação que recebo do meu desempenho me fornece uma boa avaliação do esforço que coloco no meu trabalho	A avaliação que recebo do meu desempenho me fornece uma boa avaliação do esforço que coloco no meu trabalho
9	JUST5	Considero que a minha chefia imediata avalia o desempenho dos funcionários de forma justa	Minha chefia imediata avalia o desempenho dos funcionários de forma justa

10	JUST6	Participo do estabelecimento dos indicadores que são utilizados para avaliar meu desempenho no trabalho	A minha participação é considerada no estabelecimento dos indicadores que são utilizados para avaliar meu desempenho no trabalho
11	JUST7	Percebo que as recompensas que recebo são baseadas no meu desempenho	As recompensas que recebo são baseadas no meu desempenho
12	JUST8	Sinto-me informado pela minha chefia imediata sobre o quão bem estou desempenhando minhas atividades	Obtenho informações sobre o quão bem estou desempenhando minhas atividades sob o ponto de vista da minha chefia
13	PREC1	Considero que a avaliação de desempenho que recebo é precisa em relação ao meu desempenho	A avaliação de desempenho que recebo é precisa em relação ao meu desempenho
14	PREC2	Considero que a avaliação do meu desempenho reflete o meu verdadeiro desempenho	A avaliação do meu desempenho reflete o meu verdadeiro desempenho
15	PREC3	Minha chefia imediata e eu atualizamos minhas metas conforme os objetivos da organização mudam	Minha chefia imediata e eu atualizamos minhas metas conforme os objetivos da organização mudam
16	UTIL1	Acredito que a avaliação de desempenho me auxilia a enxergar chances de crescimento na organização	Me ajuda a enxergar chances de crescimento na organização
17	UTIL2	Acredito que eu aprendo muito com a avaliação de desempenho que recebo	Aprendo muito com a avaliação de desempenho que recebo
18	UTIL3	Considero que o processo de avaliação do desempenho me ajuda a me desenvolver como profissional	O processo de avaliação do desempenho me ajuda a me desenvolver como profissional
19	UTIL4	Percebo que a avaliação de desempenho me fez refletir sobre o meu desempenho	Me faz refletir sobre o meu desempenho

Escala de Percepção de Aprendizagem

1	ATIT1	Considero que o processo de avaliação de desempenho é agradável	O processo de GD/AD é agradável
2	ATIT2	Vejo os benefícios que a avaliação de desempenho me traz	Me traz benefícios
3	ATIT3	Considero que participar da avaliação de desempenho é divertido	Participar da GD/AD é divertido
4	ATIT4	Participo da avaliação de desempenho para o meu próprio bem	Participo da GD/AD para o meu próprio bem
5	ATIT5	Sinto que tenho que participar da avaliação de desempenho	Devo continuar participando da GD/AD
6	ATIT6	Sinto-me bem ao participar da avaliação de desempenho	Me sinto bem ao participar da GD/AD
7	ATIT7	Sinto-me disposto(a) a participar da avaliação de desempenho por decisão pessoal	Estou disposto(a) a participar da GD/AD por decisão pessoal
8	ATIT8	Considero que participar da avaliação de desempenho é algo que eu tenho que fazer	Participar da GD/AD é algo que eu tenho que fazer
9	ATIT9*	Sinto-me disposto(a) a participar da avaliação de desempenho porque eu não tenho escolha*	Estou disposto(a) a participar da GD/AD porque eu não tenho escolha*
10	ATIT10*	Sinto-me disposto(a) a participar da avaliação de desempenho porque sou obrigado(a)*	Estou disposto(a) a participar da GD/AD porque sou obrigado(a)*

11	ATIT11*	Sinto-me disposto(a) a participar da avaliação de desempenho, mas não tenho certeza se é bom realizá-la*	Estou disposto(a) a participar da GD/AD, mas não tenho certeza se é bom realizá-la*
12	ATIT12*	Sinto-me disposto(a) a participar da avaliação de desempenho, mas não tenho certeza se vale a pena*	Estou disposto(a) a participar da GD/AD, mas não tenho certeza se vale a pena*
13	CONH1	Acredito que a avaliação de desempenho esclarece as expectativas em relação ao meu trabalho	As expectativas em relação ao meu trabalho são esclarecidas
14	CONH2	Acredito que a avaliação de desempenho fornece informações sobre minhas contribuições pessoais e meu valor na organização	A GD/AD fornece informações sobre minhas contribuições pessoais e meu valor na organização
15	CONH3	Minhas tarefas ficam claras para mim	Minhas tarefas ficam claras para mim
16	CONH4	Obtenho as habilidades necessárias para executar as tarefas relacionadas ao meu papel na avaliação de desempenho	Obtenho as habilidades necessárias para executar as tarefas relacionadas ao meu papel na GD/AD
17	CONH5	Sei o que se espera de mim em relação à execução das tarefas	Conheço melhor o que se espera de mim em relação à execução das tarefas que devo realizar
18	HABI1	Considero-me competente para executar minhas tarefas relacionadas ao meu papel na avaliação de desempenho	Posso executar minhas tarefas relacionadas ao meu papel na GD/AD
19	HABI2	Consigo receber ou fornecer informações claras e diretas sobre o desempenho esperado	Posso receber ou fornecer informações claras e diretas sobre o meu desempenho esperado
20	HABI3	Sei como receber ou fornecer feedback mesmo se não forem momentos formais de feedback ou de avaliação do desempenho	Sei como receber ou fornecer feedback mesmo se não forem momentos formais de devolutiva ou avaliativos
21	HABI4	Sinto-me capaz de fazer mudanças em meu desempenho geral do trabalho baseadas na minha avaliação	Sou capaz de fazer mudanças em meu desempenho a partir da avaliação que recebo
22	HABI5	Sinto-me capaz de receber ou fornecer feedback frequente e contínuo sobre o trabalho realizado	Posso tranquilamente receber ou fornecer feedback sobre o trabalho realizado
23	HABI6	Sinto-me treinado para executar minhas atribuições relacionadas ao meu papel na avaliação de desempenho	Me sinto suficientemente capaz de desempenhar meu papel como avaliado(a) ou avaliador(a) na GD/AD

Escala de Percepção de Transferência

1	ATRA1	Acredito que o tempo voa quando estou trabalhando	O tempo voa quando estou trabalhando
2	ATRA2	Considero que o meu trabalho é útil	Meu trabalho é útil
3	ATRA4	Esqueço tudo o que se passa a minha volta quando estou trabalhando	Esqueço tudo o que se passa ao meu redor quando estou trabalhando
4	ATRA5	Estou bastante satisfeito(a) com o meu trabalho	Estou satisfeito(a) com o meu trabalho
5	ATRA6	Estou imerso(a) no meu trabalho	Estou imerso(a) no meu trabalho
6	ATRA7	Estou interessado(a) no meu emprego atual	Estou interessado(a) no meu emprego atual
7	ATRA8	Estou orgulhoso(a) do que faço neste trabalho	Estou orgulhoso(a) do que faço neste trabalho

8	ATRA9	Gosto de conversar sobre a minha organização com pessoas de fora dela	Gosto de conversar sobre a minha organização com pessoas de fora dela
9	ATRA10	Gosto de trabalhar aqui	Gosto de trabalhar na minha organização
10	ATRA11	Gosto muito do meu trabalho	Gosto muito do meu trabalho
11	ATRA12	Sinto que “deixo-me ir” quando estou trabalhando	“Me deixo ir” quando estou trabalhando
12	ATRA13	Sinto que a organização representa um grande significado pessoal para mim	A organização tem um grande significado pessoal para mim
13	ATRA14	Sinto que meu trabalho vale a pena	Meu trabalho vale a pena
14	BEME1	Sinto sensações positivas a partir do meu trabalho	Meu trabalho gera sensações positivas
15	DESE1	Considero a qualidade do meu trabalho excelente	A qualidade do meu trabalho é excelente
16	DESE3	Empenho-me na realização do meu trabalho	Me empenho na realização do meu trabalho
17	DESE4	Faço um esforço extra na realização do meu trabalho	Faço um esforço extra na realização do meu trabalho
18	DESE6	Tenho um desempenho melhor do que o esperado	Meu desempenho está melhor do que o esperado
19	DESE7	Tenho um desempenho melhor do que o que pode ser caracterizado como desempenho aceitável	Meu desempenho está melhor do que o que é considerado um desempenho aceitável
20	DESE8	Tento trabalhar o máximo possível	Tento trabalhar o máximo possível
21	MOTI1	Considero meu trabalho uma motivação em si	Meu trabalho é uma motivação por si só
22	MOTI3	Considero que este trabalho preenche meus planos de carreira	Este trabalho atende aos meus planos de carreira
23	MOTI4	Considero que o meu trabalho me motiva	Meu trabalho me motiva
24	RELA3	Sou auxiliado(a) em relação a obter os recursos de que preciso para fazer meu trabalho	Meus colegas e chefia me auxiliam no que preciso para fazer meu trabalho
25	RELA4	Tenho clareza sobre o quão aceitável meu trabalho é para o meu chefe	Obtenho clareza sobre o quão aceitável meu trabalho é para a minha chefia

Nota: itens com * são os itens com avaliação reversa. Legenda: AFET: Afetividade, JUST: Justiça, PREC: Precisão, UTIL: Utilidade, ATIT: Atitude, CONH: Conhecimento, HABI: Habilidade, ATRA: Atitude no trabalho, BEME: Bem-estar, DESE: Desempenho no trabalho, MOTI: Motivação, RELA: Relações interpessoais.

Com os itens validados teórica e semanticamente, iniciou-se o processo de validação empírica das medidas de Reação, Aprendizagem e Transferência. Esse processo envolveu a coleta de dados e a análise fatorial exploratória e pode ser conferido na subseção a seguir, na Fase B, de validação empírica.

4.2. Fase B: Validação empírica

4.2.1. Subfase 1B: Análise Fatorial Exploratória (AFE)

A análise fatorial exploratória (AFE) foi realizada com o objetivo de verificar a adequação dos dados da amostra ao método, permitindo assim a definição empírica dos fatores para as escalas de Reação, Aprendizagem e Transferência.

O método de extração por eixos principais (*Principal Axis Factoring* – PAF) é empregado na análise fatorial para identificar e validar estruturas fatoriais (Tabachnick & Fidell, 2007). Existem dois problemas a serem evitados nesse tipo de análise: a) correlações que não são suficientemente elevadas; e b) correlações excessivamente altas. Recomenda-se eliminar variáveis com correlações inferiores a 0,3, assim como correlações extremamente altas podem causar multicolinearidade nos dados, dificultando a identificação da contribuição individual de cada variável para um fator. Assim, também é necessário avaliar as variáveis com correlações fortes ($r > 0,8$) e considerar sua exclusão, se necessário (Field, 2020).

Diante disso, foi feita uma matriz de correlações (matriz *R*) para cada uma das escalas, o que justificou, conforme regras apontadas anteriormente, a retirada de itens antes da extração por eixos principais, contendo também as razões para tais decisões.

No caso das escalas de Reação e Aprendizagem, apenas um item de cada uma apresentou níveis de correlação que poderiam justificar a sua exclusão. Entretanto, em uma análise aprofundada desses itens, foram considerados importantes por representarem comportamentos previstos na teoria e que não estavam abarcados por outros itens. Assim, optou-se por incluir todos os itens inicialmente previstos na análise subsequente. Já no caso da escala de Transferência, quatro itens apresentaram, conforme Tabela 32.

Tabela 32

Variáveis retiradas a partir da matriz de correlações

Escala	Variável - Item	Procedimento	Justificativa
Transferência	ATRA4 - Esqueço tudo o que se passa ao meu redor quando estou trabalhando	Variável retirada da AFE	Retirada por conter baixas correlações ($r < 0,3$) com ATRA9, ATRA10 e ATRA13
Transferência	DESE3 - Me empenho na realização do meu trabalho	Variável retirada da AFE	Retirada por conter baixas correlações ($r < 0,3$) com DESE4, DESE6 e DESE7
Transferência	DESE4 - Faço um esforço extra na realização do meu trabalho	Variável retirada da AFE	Retirada por conter baixas correlações ($r <$

			0,3) com DESE5, DESE6, DESE7 e DESE8
Transferência	DESE6 - Tento trabalhar o máximo possível	Variável retirada da AFE	Retirada por conter baixas correlações ($r < 0,3$) com DESE3, DESE4, DESE5, DESE6 e DESE7

Apenas itens da escala de Transferência foram retirados em razão da sua correlação com outros. As variáveis retiradas apresentavam baixa correlação com outras variáveis do mesmo fator. A exclusão dessas variáveis é pertinente, uma vez que, por terem baixa correlação, podem representar ideias muito distantes, ou até mesmo opostas, o que poderia vir a comprometer o poder explicativo do instrumento. Com isso, a escala de Transferência passou a contar com 21 itens.

Após a retirada dos itens, foi realizada a AFE em cada uma das escalas com o RDWS (Robust Diagonally Weighted Least Squares) como método de extração, por meio da técnica de Análise Paralela com permutação aleatória dos dados observados (Timmerman & Lorenzo-Seva, 2011) e a rotação utilizada foi a Robust Promin (Lorenzo-Seva & Ferrando, 2019c). Os resultados da AFE de cada uma das escalas são apresentados abaixo.

Escala de Percepção de Reação – AFE

A escala de Reação foi inicialmente submetida à Análise Fatorial Exploratória (AFE) com uma matriz de 19 itens, mostrando-se adequada para interpretação com base no teste de esfericidade de Bartlett significativo ($\chi^2 = 1230,0$; gl = 171; $p < 0,001$) e no índice Kaiser-Meyer-Olkin (KMO) de 0,892, considerado excelente.

Na configuração inicial, embora tenha sido simulada uma estrutura com quatro fatores, os autovalores indicaram a possibilidade de até três fatores, com o primeiro apresentando autovalor de 10,07 e o terceiro, de 1,19. No entanto, a análise paralela sugeriu a retenção de dois fatores, que explicaram 53% e 11,84% da variância, respectivamente. Apesar do autovalor representativo do terceiro fator, sua variância aleatória (9,48%) foi maior do que a variância explicada pelo fator original (6,29%), indicando tratar-se de um fator espúrio.

Com base nesses resultados, a escala de Reação foi novamente submetida à AFE, desta vez considerando dois fatores e mantendo os 19 itens. Nessa configuração, as cargas fatoriais de cada item foram analisadas para identificar possíveis cargas cruzadas. Como resultado, os itens JUST2, JUST19 e UTIL4 foram excluídos da análise devido a cargas fatoriais abaixo de 0,3 ou por apresentarem cargas cruzadas superiores a 0,3 em mais de um fator.

Após essa exclusão, a nova análise, agora da escala com 16 itens, manteve a estrutura fatorial de dois fatores estável, conforme mostrado na Tabela 33.

Tabela 33

Resultados da Análise Paralela

Variáveis	Percentual de variância explicada dos dados reais	Percentual de variância explicada dos dados aleatórios (95% IC)
1	55,1581*	152,280
2	14,1118*	131,589
3	68,343	118,329
4	48,332	106,136
5	38,764	96,045
6	30,967	87,120
7	25,855	78,629
8	24,395	70,078
9	19,375	62,957
10	19,375	55,375
11	11,872	48,426
12	0,9229	41,324
13	0,4179	34,085
14	0,2176	27,401
15	0,1056	18,295

Nota: Os fatores a serem retidos possuem um * e estão em negrito na tabela.

Portanto, o número de fatores retidos foi de dois, uma vez que apenas esses fatores nos dados reais apresentaram percentuais de variância explicada superiores aos percentuais dos dados aleatórios.

As cargas fatoriais dos itens estão detalhadas na Tabela 34, onde também são apresentados os índices de Confiabilidade Composta e a Variância Média Extraída (AVE)

(Valentini & Damásio, 2016) e as estimativas de replicabilidade dos escores fatoriais (H-index; Ferrando & Lorenzo-Seva, 2018).

Tabela 34

Estrutura fatorial da Escala de Reação

Variável	Item	Fator 1	Fator 2
AFET3	Fico feliz em receber o resultado da minha avaliação de desempenho	0,763	0,223
AFET4*	Fico tenso(a) ao ter meu desempenho avaliado*	-0,080	0,732
AFET6*	Sinto raiva ao ter meu desempenho avaliado*	0,191	0,777
JUST3	A avaliação que recebo do meu desempenho é equânime	-0,189	-0,542
JUST4	A avaliação que recebo do meu desempenho me fornece uma avaliação adequada do trabalho que eu realizo	0,715	0,240
JUST5	A avaliação que recebo do meu desempenho me fornece uma boa avaliação do esforço que coloco no meu trabalho	0,877	0,132
JUST11	Minha chefia imediata avalia o desempenho dos funcionários de forma justa	0,844	0,114
JUST15	A minha participação é considerada no estabelecimento dos indicadores que são utilizados para avaliar meu desempenho no trabalho	0,744	0,074
JUST16	As recompensas que recebo são baseadas no meu desempenho	0,770	0,007
PREC7	A avaliação de desempenho que recebo é precisa em relação ao meu desempenho	0,649	-0,034
PREC8	A avaliação do meu desempenho reflete o meu verdadeiro desempenho	0,888	0,120
PREC10	Minha chefia imediata e eu atualizamos minhas metas conforme os objetivos da organização mudam	0,707	0,103
UTIL2	Me ajuda a enxergar chances de crescimento na organização	0,721	-0,058
UTIL4	Aprendo muito com a avaliação de desempenho que recebo	0,795	-0,213

UTIL6	O processo de avaliação do desempenho me ajuda a me desenvolver como profissional	0,865	-0,290
UTIL7	Me faz refletir sobre o meu desempenho	0,817	-0,203
Confiabilidade composta ²		0,953	0,711
Variância Média Extraída (AVE)		61,53%	45,51
H-latent		0,965	0,821
H-observed		0,922	0,769

Nota: cargas de cada fator destacadas em negrito. As variáveis que apresentam * ao lado são itens reversos.

Os índices de ajuste do instrumento foram adequados ($\chi^2 = 535,92$, $gl = 185$; $p < 0,001$; RMSEA [95% IC BCa] = 0,050 [0,0000 - 0,0760]; CFI [95% IC BCa] = 0,994 [0,985 - 1,003]; TLI [95% IC BCa] = 0,992 [0,980 - 1,004]). As cargas fatoriais, que refletem a correlação entre os itens e os fatores, demonstraram valores elevados (acima de 0,7), sugerindo uma forte associação entre os itens e seus respectivos fatores.

A confiabilidade dos fatores, avaliada por meio da Confiabilidade Composta, apresentou valores acima de 0,7, indicando boa confiabilidade para ambos os fatores. O Fator 1 obteve um valor de 0,953, sugerindo excelente consistência interna, enquanto o Fator 2 obteve 0,711, considerado aceitável.

Com relação à AVE, o Fator 1 apresentou valor de 0,6153, indicando boa validade convergente, ou seja, os itens medem efetivamente o mesmo conceito. No entanto, o Fator 2 apresentou um AVE de 0,4551, ligeiramente abaixo do valor recomendado de 0,5, sugerindo que talvez a variância explicada não seja tão forte.

Por essa razão, considerou-se pertinente avaliar os indicadores de unidimensionalidade da escala, a fim de avaliar se havia elementos suficientes para desconsiderar o Fator 2 e considerar a escala com apenas uma dimensão. Para isso, foram avaliados os índices de avaliação da proximidade à unidimensionalidade: a) *Unidimensional Congruence* (UniCo); b) *Explained Common Variance* (ECV); e c) *Mean of Item Residual Absolute Loadings* (MIREAL) (Ferrando & Lorenzo-Seva, 2018). A

² O cálculo da confiabilidade composta e a variância média extraída foi realizado pelo site: <https://vecon.ppgmcs.com.br/>

Tabela 35 apresenta esses índices juntamente aos seus respectivos intervalos de confiança, evidenciando a estrutura unidimensional.

Tabela 35

Avaliação de Aproximação à Unidimensionalidade da Escala de Reação

Avaliação de Aproximação à Unidimensionalidade	Índice	Avaliação geral	Avaliação geral (95% IC)	Interpretação
UniCo	0.828	(0.806 - 0.868)	< 0,95	Inadequado
ECV	0.806	(0.733 - 0.848)	> 0,85	Inadequado
MIREAL	0.229	(0.178 - 0.264)	< 0,30	Adequado

De modo geral, a escala não demonstrou características essencialmente unidimensionais, apresentando apenas o índice MIREAL abaixo de 0,30, conforme os parâmetros recomendados pela literatura. Assim, considerando que a solução bidimensional se mostrou clara, replicável e resultou em estimativas de pontuação fatoriais precisas e confiáveis para ambos os fatores, além de convergir teoricamente com a literatura que avalia a reação modelada com fatores distintos (*e.g.* Keeping & Levy, 2000; Pichler, 2012; Sharma et al., 2016), não foi considerado relevante tratar a escala como unidimensional.

Quanto à AVE do Fator 2, que inicialmente motivou a investigação sobre a unidimensionalidade, recomenda-se que seja revisitada em estudos futuros, com possíveis ajustes nos itens existentes ou a inclusão de novos itens para fortalecer a validade convergente.

Por fim, a medida de replicabilidade da estrutura fatorial, avaliada pelos índices *H-latent* e *H-observed*, também foi satisfatória. O Fator 1 apresentou valores altos e próximos (*H-latent* = 0,965; *H-observed* = 0,922), indicando que a confiabilidade do construto latente é consistentemente refletida nos itens observados. Para o Fator 2, os índices foram *H-latent* = 0,821 e *H-observed* = 0,769, mostrando uma boa consistência, embora com uma leve diferença que sugere menor estabilidade em comparação ao Fator 1. Apesar disso, o fator ainda possui uma estrutura suficientemente forte para ser considerado confiável.

Assim, foi adotada a estrutura bidimensional, em que o Fator 1 é caracterizado por variáveis com cargas fatoriais elevadas, acima de 0,7, relacionadas à percepção de justiça, utilidade e precisão da avaliação de desempenho. Esses critérios, segundo a teoria, são aglutinados no conceito de **Reações Cognitivas**, pois refletem como as pessoas pensam acerca do sistema de GD (Schleicher et al., 2019). Embora o Fator 1 inclua um item associado à afetividade – especificamente, relacionado à felicidade com o sistema de GD –, esse item pode ser interpretado de forma consciente com o domínio cognitivo. A compreensão de um sentimento positivo, como a felicidade em relação ao sistema, pode ser mediada por um processo cognitivo, no qual as pessoas primeiro analisam e refletem sobre o funcionamento do sistema e, em seguida, interpretam essas percepções como atributos positivos, capazes de gerar satisfação ou bem-estar. Dessa forma, mesmo no caso de itens aparentemente afetivos, estes podem ser plausivelmente enquadrados como parte das reações cognitivas, reforçando a coerência do item AFET3 dentro da estrutura fatorial adotada.

O Fator 2 agrupa itens relacionados a aspectos emocionais e sentimentais, com ênfase em reações negativas ao processo de avaliação de desempenho, como ansiedade e raiva (AFET4* e AFET6*), itens que apresentaram cargas fatoriais elevadas (0,732 e 0,777). Esses itens, formulados como reversos, captam respostas opostas ao comportamento esperado nos demais itens.

O Fator 2 também inclui a variável JUST3, relacionada à equanimidade na avaliação de desempenho, que apresentou carga fatorial negativa. Isso sugere que respostas altas nesse item refletem a ausência de reações emocionais negativas ou um estado de neutralidade emocional. O modelo aponta que o Fator 2 reflete um *continuum* de Reações Afetivas, variando entre a presença de emoções negativas e a ausência dessas reações, como quando há a equanimidade. Assim, o fator engloba tanto a experiência de sentimentos negativos quanto a percepção de sua ausência, configurando um aspecto emocional. Esses resultados se alinham ao conceito de Reações Afetivas descrito por Brett e Atwater (2001), Keeping e Levy (2000) e Schleicher et al. (2019), que associam essas reações à forma como as pessoas sentem o sistema de GD.

Portanto, a Escala de Percepção de Reação foi definida como composta por dois fatores: **Reações Cognitivas**, com 13 itens, e **Reações Afetivas**, com 3 itens. Estatisticamente, o modelo com dois fatores apresentou ajustes adequados e confiáveis, evidenciados pelos índices de ajuste, confiabilidade composta e variância explicada. A adoção de dois fatores foi corroborada em partes pela análise teórica, mas adequada pela

estatística, destacando-se como a solução mais robusta para mensurar as percepções de reação no contexto da GD. Essa estrutura, embora diferente da originalmente proposta com cinco fatores, mostrou-se mais parcimoniosa e consistente, eliminando redundâncias e itens pouco representativos identificados durante o processo de validação.

Escala de Percepção de Aprendizagem - AFE

A segunda escala analisada na AFE foi a de Aprendizagem, composta por 23 itens e com três fatores. Os resultados indicaram o teste de esfericidade de Bartlett's significativo (1226,1, $gl = 253$, $p < 0,001$), sugerindo que as correlações entre as variáveis são suficientemente altas para justificar a AFE. O KMO foi de 0,87, o que indica uma boa adequação dos dados para a análise fatorial.

Com essa configuração, a Análise Paralela sugeriu a retenção de dois fatores principais, pois os dois primeiros fatores reais (com variância explicada de 43.66% e 11.83%) estavam acima das médias aleatórias dos dados simulados.

Assim, a escala de Aprendizagem foi novamente submetida à AFE, agora com dois fatores, ainda com os 23 itens. Nessa configuração, foi analisada a carga fatorial de cada um dos itens a fim de identificar cargas cruzadas ou baixas, menores de 0,7, razão pela qual os itens ATIT 4, ATIT8, ATIT10*, ATIT11* foram inicialmente retirados da análise. Nesse ponto, a análise paralela, agora com 19 itens, sugeriu apenas uma dimensão (com variância explicada de 49,41%).

Na nova análise realizada com apenas um fator, ainda se mostrou necessário excluir os itens com cargas fatoriais baixas, como o ATIT2, ATIT5, ATIT7, ATIT9, ATIT12, HABI3, HABI5 e HABI6. Assim, a escala foi finalizada com 11 itens e um fator e submetida à análise uma última vez.

A estrutura unidimensional do fator mostrou-se consistente, e os resultados da Análise Paralela foram corroborados pelos índices de avaliação da proximidade à unidimensionalidade: a) *Unidimensional Congruence* (UniCo); b) *Explained Common Variance* (ECV); e c) *Mean of Item Residual Absolute Loadings* (MIREAL) (Ferrando & Lorenzo-Seva, 2018). A Tabela 36 apresenta esses índices juntamente dos seus respectivos intervalos de confiança, evidenciando a estabilidade da estrutura unidimensional.

Tabela 36

Avaliação de Aproximação à Unidimensionalidade

Avaliação de Aproximação à Unidimensionalidade	Índice	Avaliação geral	Avaliação geral (95% IC)	Interpretação
UniCo	0,987	(0,981 - 0,995)	> 0,95	Adequado
ECV	0,897	(0,874 - 0,933)	> 0,85	Adequado
MIREAL	0,224	(0,158 - 0,265)	< 0,30	Adequado

Todos os índices sugeriram que a medida deve ser interpretada como unidimensional. Assim, optou-se por manter a análise com um único fator. A variância explicada desse modelo ficou em 66,58% e melhorou a estrutura fatorial. As cargas fatoriais dos itens estão detalhadas na Tabela 37, sendo também fornecidos os índices de Confiabilidade Composta (Valentini; Damásio, 2016), juntamente das estimativas de replicabilidade dos escores fatoriais (H-index; Ferrando & Lorenzo-Seva, 2018).

Tabela 37

Estrutura fatorial da Escala de Aprendizagem

Variável	Item	Fator 1	Comunalidade
ATIT1	O processo de GD/AD é agradável	0,744	0,553
ATIT3	Participar da GD/AD é divertido	0,664	0,441
ATIT6	Me sinto bem ao participar da GD/AD	0,802	0,643
CONH1	As expectativas em relação ao meu trabalho são esclarecidas	0,803	0,645
CONH2	A GD/AD fornece informações sobre minhas contribuições pessoais e meu valor na organização	0,713	0,508
CONH3	Minhas tarefas ficam claras para mim	0,842	0,709
CONH4	Obtenho as habilidades necessárias para executar as tarefas relacionadas ao meu papel na GD/AD	0,672	0,451
CONH5	Conheço melhor o que se espera de mim em relação à execução das tarefas que devo realizar	0,827	0,684

HABI1	Posso executar minhas tarefas relacionadas ao meu papel na GD/AD	0,756	0,572
HABI2	Posso receber ou fornecer informações claras e diretas sobre o meu desempenho esperado	0,827	0,685
HABI4	Sou capaz de fazer mudanças em meu desempenho a partir da avaliação que recebo	0,654	0,427
Confiabilidade composta ³		0,9364	
Variância Média Extraída (AVE)		0,5744	
H-latent		0,942	
H-observed		0,886	

Os resultados indicaram o teste de esfericidade de Bartlett's significativo (938,3, $gl = 55$, $p < 0,001$), sugerindo que as correlações entre as variáveis são suficientemente altas para justificar a AFE. O KMO foi de 0,90, o que indica uma adequação muito boa dos dados para a análise fatorial.

Os índices de ajuste do instrumento foram adequados ($\chi^2 = 31,54$, $gl = 54$, $\chi^2/gl = 0,58$; RMSEA [95% IC BCa] = 0,048 [0,0000 - 0,0938]; TLI [95% IC BCa] = 0,989 [0,951 - 1,042]; CFI [95% IC BCa] = 0,991 [0,961 - 1,034]). As cargas fatoriais indicam a correlação entre os itens e o fator. Valores altos (acima de 0,7) indicam uma forte associação entre o item e o fator.

A confiabilidade composta é alta (0,9364), o que demonstra excelente consistência interna dos itens. A AVE de 0,5744 está acima do limite mínimo de 0,50, sugerindo que o fator único explica uma quantidade substancial da variância total dos itens.

A replicabilidade da estrutura fatorial, com H-*latent* de 0,942, é alta, indicando que o fator latente explica bem a variância dos itens. Valores acima de 0,70 são considerados bons, e acima de 0,90, como neste caso, excelentes. O H-*observed* de 0,886 também é elevado, confirmando boa replicabilidade nos dados observados. Esses resultados sugerem alta replicabilidade do fator, tanto no nível latente quanto no observado, com intervalos de confiança reforçando a robustez da solução fatorial.

³ O cálculo da confiabilidade composta e a variância média extraída foi realizado pelo site: <https://vecon.ppgmcs.com.br/>

Estatisticamente, o modelo unifatorial apresentou ajustes adequados e confiáveis, evidenciados pelos índices de ajuste, confiabilidade composta e variância explicada. A transição de um modelo teórico multifatorial para um modelo unifatorial foi suportada por evidências empíricas que demonstram itens da escala; embora inicialmente distribuídos em múltiplos fatores, estão altamente correlacionados em um constructo único. Além disso, a confiabilidade composta e a replicabilidade latente confirmam a adequação da solução unidimensional.

Do ponto de vista teórico, essa simplificação não compromete a essência da escala, considerando a recente incorporação desse construto no campo de estudo da efetividade dos sistemas de gestão de desempenho e as idiossincrasias identificadas nas operacionalizações abordadas na revisão de literatura deste estudo, como também apontado por Schleicher et al. (2019). Pelo contrário, ela reforça que o conceito central de percepção de aprendizagem é suficientemente abrangente para integrar aspectos que anteriormente eram tratados de forma separada. A solução final representa uma convergência coerente entre fundamentação teórica e evidências estatísticas, proporcionando uma abordagem mais parcimoniosa do instrumento.

Escala de Percepção de Transferência - AFE

A terceira escala analisada na AFE foi a de Transferência, composta por 21 itens e com cinco fatores. Os resultados indicaram o teste de esfericidade de Bartlett's significativo (1195,4, gl = 561, $p < 0,001$) e do KMO de 0,88, considerando excelente. Esses resultados indicaram que a matriz de correlação dos itens era interpretável. Com essa configuração, a Análise Paralela sugeriu uma escala unifatorial, ou seja, com apenas um fator sendo o mais representativo, com 58% de variância no primeiro fator, conforme Tabela 38.

Tabela 38

Resultados da Análise Paralela com cinco fatores

Variáveis	Percentual de variância explicada dos dados reais	Percentual de variância explicada dos dados aleatórios (95% IC)
1	58.5281*	118.137
2	72.998	105.356

3	63.376	95.320
4	46.697	87.520
5	39.091	80,630
6	28.627	74.651
7	24.605	68.628
8	22.498	63.066
9	20,412	58.485
10	17.863	53.102
11	15.394	48.652
12	13.796	44.065
13	13.031	40,020
14	0,9395	36.199
15	0,8186	31.988
16	0,6357	27.641
17	0,5366	23.587
18	0,3353	19.289
19	0,2333	15.091
20	0,1341	0,9800

Nota: O fator a ser retido possui um * e está em negrito na tabela.

Assim, a escala de Transferência foi novamente submetida à AFE, agora com estrutura fatorial com apenas um fator, mas ainda com os 21 itens. Os índices de ajuste se mostraram adequados ($\chi^2 = 164,59$, gl = 189, $\chi^2/\text{gl} = 0,8$; RMSEA [95% IC BCa] = 0,000 [0,0000 - 0,0678]; TLI [95% IC BCa] = 1.005 [0,992 - 1.050]; CFI [95% IC BCa] = 0,999 [0,967 - 1.030]). Com essa estrutura, não se mostrou necessária a exclusão de nenhum item com carga cruzada. Os resultados estão na Tabela 39.

Tabela 39

Estrutura fatorial da Escala de Transferência

Variável	Item	Fator 1	Comunalidade
ATRA1	O tempo voa quando estou trabalhando	0,727	0,529
ATRA2	Meu trabalho é útil	0,785	0,616
ATRA5	Estou satisfeito(a) com o meu trabalho	0,848	0,720
ATRA6	Estou imerso(a) no meu trabalho	0,752	0,566

ATRA7	Estou interessado(a) no meu emprego atual	0,873	0,762
ATRA8	Estou orgulhoso(a) do que faço neste trabalho	0,851	0,724
ATRA9	Gosto de conversar sobre a minha organização com pessoas de fora dela	0,713	0,509
ATRA10	Gosto de trabalhar na minha organização	0,782	0,611
ATRA11	Gosto muito do meu trabalho	0,913	0,833
ATRA12	"Me deixo ir" quando estou trabalhando	0,706	0,499
ATRA13	A organização tem um grande significado pessoal para mim	0,741	0,549
ATRA14	Meu trabalho vale a pena	0,820	0,672
BEME1	Meu trabalho gera sensações positivas	0,818	0,670
DESE1	A qualidade do meu trabalho é excelente	0,675	0,456
DESE6	Meu desempenho está melhor do que o esperado	0,550	0,302
DESE7	Meu desempenho está melhor do que o que é considerado um desempenho aceitável	0,539	0,290
MOTI1	Meu trabalho é uma motivação por si só	0,797	0,635
MOTI3	Este trabalho atende aos meus planos de carreira	0,753	0,568
MOTI4	Meu trabalho me motiva	0,896	0,802
RELA3	Meus colegas e chefia me auxiliam no que preciso para fazer meu trabalho	0,495	0,245
RELA4	Obtenho clareza sobre o quão aceitável meu trabalho é para a minha chefia	0,497	0,247

Nota: Cargas fatoriais < 0,3 destacadas em negrito.

Com base na análise das cargas fatoriais e comunalidades, quatro variáveis foram retiradas da análise devido à sua baixa representatividade no modelo: DESE6, DESE7, RELA3 e RELA4. A exclusão desses itens refletiu uma melhora no ajuste geral do modelo e a consistência da escala. Os índices de ajuste se mostraram adequados ($\chi^2 = 150,63$, gl = 119, $\chi^2/\text{gl} = 1,26$; RMSEA [95% IC BCa] = 0,048 [0,0000 - 0,0685]; TLI [95% IC BCa] = 0,994 [0,987 - 1,003]; CFI [95% IC BCa] = 0,995 [0,989 - 1,003]).

A estrutura unidimensional do fator mostrou-se consistente, e os resultados da Análise Paralela foram corroborados pelos índices de avaliação da proximidade à unidimensionalidade: a) *Unidimensional Congruence* (UniCo); b) *Explained Common Variance* (ECV); e c) *Mean of Item Residual Absolute Loadings* (MIREAL) (Ferrando & Lorenzo-Seva, 2018). A Tabela 40 apresenta esses índices juntamente dos seus respectivos intervalos de confiança, evidenciando a estabilidade da estrutura unidimensional.

Tabela 40

Avaliação de Aproximação à Unidimensionalidade

Aproximação à Unidimensionalidade	Índice	Avaliação geral	Avaliação geral (95% IC)	Interpretação
UniCo	0.964	(0.894 - 0.988)	> 0,95	Adequado
ECV	0.891	(0.831 - 0.928)	> 0,85	Adequado
MIREAL	0.213	(0.182 - 0.279)	< 0,30	Adequado

Todos os índices sugeriram que a medida deve ser interpretada como unidimensional. Assim, optou-se por manter a análise com um único fator, mas agora com os 17 itens. As cargas fatoriais dos itens estão detalhadas na Tabela 41, sendo também fornecidos os índices de Confiabilidade Composta (Valentini & Damásio, 2016), juntamente das estimativas de replicabilidade dos escores fatoriais (H-indez; Ferrando & Lorenzo-Seva, 2018).

Tabela 41

Cargas fatoriais

Variável	Item	Fator 1
ATRA1	O tempo voa quando estou trabalhando	0,717
ATRA2	Meu trabalho é útil	0,746
ATRA5	Estou satisfeito(a) com o meu trabalho	0,848
ATRA6	Estou imerso(a) no meu trabalho	0,728
ATRA7	Estou interessado(a) no meu emprego atual	0,868
ATRA8	Estou orgulhoso(a) do que faço neste trabalho	0,828

ATRA9	Gosto de conversar sobre a minha organização com pessoas de fora dela	0,706
ATRA10	Gosto de trabalhar na minha organização	0,807
ATRA11	Gosto muito do meu trabalho	0,926
ATRA12	"Me deixo ir" quando estou trabalhando	0,687
ATRA13	A organização tem um grande significado pessoal para mim	0,751
ATRA14	Meu trabalho vale a pena	0,796
BEME1	Meu trabalho gera sensações positivas	0,790
DESE1	A qualidade do meu trabalho é excelente	0,627
MOTI1	Meu trabalho é uma motivação por si só	0,796
MOTI3	Este trabalho atende aos meus planos de carreira	0,723
MOTI4	Meu trabalho me motiva	0,901
Confiabilidade composta ⁴		0,9639
Variância Média Extraída (AVE)		0,6130
H-latent		0,971
H-observed		0,903

Os índices, de maneira geral, mantiveram-se consistentes e refletiram as expectativas: a confiabilidade e a replicabilidade aumentaram, enquanto os resíduos também cresceram devido à redução no número de fatores e à menor variância explicada pelo modelo. Ambos os modelos apresentaram excelente ajuste, com o primeiro destacando-se levemente nos índices de ajuste absoluto. No entanto, o segundo modelo ainda permanece dentro dos parâmetros excelentes de ajuste (Brown, 2015).

Os resultados indicam que a maioria dos itens apresenta cargas fatoriais superiores a 0,7, com destaque para ATRA11 ("Gosto muito do meu trabalho") e ATRA1 ("O tempo voa quando estou trabalhando"), que apresentaram as maiores cargas, 0,926 e 0,717, respectivamente. Apenas dois itens, ATRA12 ("Me desligo...") e DESE1 ("A qualidade

⁴ O cálculo da confiabilidade composta e a variância média extraída foi realizado pelo site: <https://vecon.ppgmcs.com.br/>

do meu trabalho...”), apresentaram cargas fatoriais abaixo de 0,7, com valores de 0,687 e 0,627, embora ainda dentro do intervalo aceitável para análise.

A confiabilidade composta dos fatores foi avaliada, resultando em um valor de 0,9639, indicando uma excelente consistência interna entre os itens que compõem os fatores. Esse valor é significativamente superior ao limite recomendado de 0,7, o que sugere que os itens medem de forma confiável os construtos latentes.

Além disso, a AVE foi calculada, resultando em 0,6130, que está acima do valor mínimo de 0,5, confirmando a validade convergente dos fatores. Os índices de validade discriminante, *H-latent* (0,928) e *H-observed* (0,903), também indicam que os itens se correlacionam adequadamente com os fatores latentes, reforçando a distinção entre os diferentes construtos avaliados.

A estrutura teórica inicial para a escala de Transferência indicava cinco fatores, alinhada à suposição de que a transferência seria multidimensional, abrangendo diferentes aspectos de desempenho, motivação e percepção no trabalho. Contudo, os dados empíricos demonstraram uma forte predominância de um único fator, com o primeiro fator explicando 58% da variância total, enquanto os demais fatores apresentaram variâncias não expressivas em comparação. Os demais índices de ajuste, a análise unidimensional, a confiabilidade composta e a AVE também apresentaram valores adequados. Portanto, estatisticamente ficou evidente que a escala é melhor representada como unifatorial.

A simplificação estatística para um fator reflete as inter-relações substanciais entre os itens, captando um constructo abrangente de percepção de transferência. A solução final proporciona uma escala mais parcimoniosa, significativa e alinhada com os achados empíricos, facilitando sua aplicação e interpretação no contexto da avaliação de transferência nos sistemas de GD.

As escalas validadas de forma exploratória foram submetidas à Análise Fatorial Confirmatória (AFC) e os resultados foram apresentados na subfase 2B, na subseção seguinte.

4.2.2. Subfase 2B: Análise Fatorial Confirmatória (AFC)

A Análise Fatorial Confirmatória foi realizada com cada uma das três escalas: Escala de Percepção de Reação, com 16 itens, Escala de Percepção de Aprendizagem, com 11 itens, e Escala de Percepção de Transferência, com 17 itens após a validação da

AFE na subfase anterior. A seguir, é possível conferir os resultados da validação de cada uma das escalas nas próximas subseções.

Escala de Percepção de Reação - AFC

A estrutura com dois fatores foi avaliada utilizando o método de extração RDWS (*Robust Diagonally Weighted Least Squares*). A estrutura bidimensional apresentou resultados de ajuste contraditórios. Conforme pode ser visto na Tabela 42, os valores de qui-quadrado foram significativos e a razão qui-quadrado por graus de liberdade também foi elevada (9,587). Os índices de CFI e TLI suportaram o modelo. Entretanto, os índices de RMSEA e SRMR foram acima do aceitável.

Tabela 42

Índices de ajuste do modelo bifatorial da Escala de Reação

Escala e modelo	χ^2 (gl)	χ^2 /gl	CFI	TLI	SRMR	RMSEA (90% IC)
Escala de Reação - Original	356,908* (103)	3,465	0,989	0,987	0,112	0,131 (0,116 – 0,146)

Nota. * $p < 0,001$

Com o objetivo de melhor inspecionar os resultados, foram avaliados os índices de modificação, que apresentaram elevada covariância residual entre alguns pares de itens, em especial entre os itens UTIL6 e UTIL 7 (Índice de Modificação = 88,538) e JUST5 e JUST11 (Índice de Modificação = 68,729). Essa variância era esperada, uma vez que os itens pertencem ao mesmo construto inicialmente previsto pela própria teoria.

Com a inclusão dessa covariância residual no modelo, todos os índices de ajuste melhoraram, incluindo o RMSEA e o SRMR, com os maiores valores residuais não ultrapassando 1,00 (Tabela 43). Embora o RMSEA esteja ligeiramente acima do recomendado na literatura, o valor de 0,083 é considerado aceitável. Kenny et al. (2015) indicam que, em modelos com amostras pequenas e nível de liberdade próximo ao deste estudo, há uma chance de 16% de rejeitar incorretamente a hipótese nula para um valor de 0,05. Portanto, o RMSEA de 0,083 é interpretado como adequado neste contexto, considerando o tamanho da amostra e os graus de liberdade.

Tabela 43

Índices de ajuste do modelo bifatorial re-especificado da Escala de Reação

Escala e modelo	χ^2 (gl)	χ^2 / gl	CFI	TLI	SRMR	RMSEA (90% IC)
Escala de Reação - Re-especificado	200.177* (100)	2.001	0,996	0,995	<0,100	0,083 (0,067 – <0,100)

Nota. * $p < 0,001$

Optou-se por tratar esse modelo como uma estrutura de dois fatores que refletem a dimensão Reação, uma abordagem conceitualmente mais parcimoniosa em comparação à representação desses fatores como entidades correlacionadas. Além disso, o modelo de segunda ordem apresenta uma fundamentação teórica sólida, o que constitui um aspecto crucial na interpretação de modelos (MacCallum, 1995). Assim, a Tabela 44 demonstra as novas cargas fatoriais dos itens com a estrutura de segunda-ordem.

Tabela 44

Cargas fatoriais de Reação de segunda-ordem com 2 fatores

Fator	Variável	Estimativa	Desvio Padrão	Valor z	p	95% IC	
						Limite inferior	Limite superior
Reações Afetivas	AFET4*	0,944	0,072	13,188	< ,001	0,804	1,084
	AFET6*	0,880	0,070	12,632	< ,001	0,742	1,018
	JUST3	0,583	0,079	7,397	< ,001	0,429	0,738
	AFET3	0,730	0,045	16,319	< ,001	0,641	0,817
	JUST4	0,814	0,034	24,030	< ,001	0,748	0,881
	JUST5	0,857	0,037	23,294	< ,001	0,792	0,909
	JUST11	0,830	0,030	27,369	< ,001	0,771	0,890
Reações Cognitivas	JUST15	0,785	0,035	23,053	< ,001	0,771	0,869
	JUST16	0,755	0,037	21,117	< ,001	0,712	0,858
	PREC7	0,845	0,041	18,598	< ,001	0,675	0,834
	PREC8	0,793	0,029	28,913	< ,001	0,788	0,902
	PREC10	0,803	0,036	22,254	< ,001	0,723	0,863
	UTIL2	0,906	0,036	22,526	< ,001	0,733	0,873
	UTIL4	0,868	0,021	42,853	< ,001	0,818	0,947
	UTIL6	0,735	0,026	33,967	< ,001	0,804	0,918
	UTIL7	0,835	0,042	17,676	< ,001	0,654	0,817

Nota: Os valores da estimativa padronizada são os mesmos da coluna de estimativas da tabela.

As cargas fatoriais foram consistentes e significativas para a maioria dos itens nos fatores. Itens como AFET4 e AFET6 apresentam cargas fatoriais de 0,944 e 0,880, respectivamente, indicando que esses itens se relacionam fortemente com o fator Reações Afetivas. Itens como JUST5 e JUST11 também mostraram altas cargas fatoriais (0,857 e 0,830), reforçando a consistência interna desse fator.

Os construtos de Reação se transformaram em indicadores latentes de um fator de ordem superior. O fator de ordem superior (Reação) explica 46,8% da variância do fator Reações Afetivas (Afetivo) e 20,2% da variância do fator Reações Cognitivas (Cognitivo).

O fator de ordem superior (Reação) influencia mais as Reações Afetivas do que as Reações Cognitivas, sugerindo que a dimensão afetiva é mais sensível ou correlacionada à percepção global de reação, enquanto a dimensão cognitiva pode ser influenciada por outros aspectos específicos da experiência dos participantes.

A análise fatorial confirmatória, com ajustes e reespecificação, validou a estrutura bidimensional da Escala de Reação. O modelo de segunda-ordem proporcionou uma visão abrangente, destacando a complementaridade entre Reações Afetivas e Cognitivas. Os autores Keeping e Levy (2000) encontraram resultado semelhante em seu estudo em que a reação se ajustava melhor a um modelo no qual as suas construções são modeladas como distintas, mas relacionadas através de um fator de ordem superior. Esses achados sustentam a aplicabilidade da escala como ferramenta para avaliar a efetividade do sistema de gestão de desempenho no setor público.

Escala de Percepção de Aprendizagem - AFC

A estrutura unidimensional apresentou resultados de ajuste satisfatórios. Conforme pode ser visto na Tabela 45, os valores de qui-quadrado não foram significativos e razão qui-quadrado por graus de liberdade foi satisfatória. Os índices de CFI, TLI, SRMR e RMSEA suportaram o modelo.

Tabela 45

Índices de ajuste do modelo unifatorial de Aprendizagem

Escala	χ^2 (gl)	χ^2/gl	CFI	TLI	SRMR	RMSEA (90% IC)
Escala de Aprendizagem	62,208* (42)	1,482	0,998	0,998	0,054	0,058 (0,022 – 0,087)

Nota: * $p = 0,023$

A Tabela 46 apresenta as cargas fatoriais dos itens da escala de Aprendizagem.

Tabela 46

Cargas fatoriais de Aprendizagem

Fator	Variável	Estimativa	Desvio Padrão	Valor z	p	95% IC	
						Limite inferior	Limite superior
Aprendizagem	ATIT1	0,812	0,030	27,099	<,001	0,754	0, 871
	ATIT3	0,860	0,058	11,710	<,001	0,549	0, 770
	ATIT6	0,752	0,040	18,620	<,001	0,672	0, 831
	CONH1	0,829	0,028	29,314	<,001	0,773	0, 884
	CONH2	0,892	0,020	44,120	<,001	0,852	0, 933
	CONH3	0,784	0,036	21,242	<,001	0,713	0, 854
	CONH4	0,906	0,020	45,261	<,001	0,857	0, 934
	CONH5	0,916	0,019	48,708	<,001	0,879	0, 953
	HABI1	0,786	0,031	25,711	<,001	0,735	0, 836
	HABI2	0,833	0,027	30,814	<,001	0,780	0, 886
	HABI4	0,654	0,051	12,825	<,001	0,554	0, 753

Nota: Os valores da estimativa padronizada são os mesmos das estimativas da tabela.

Por fim, os resultados da análise fatorial confirmatória demonstram alinhamento consistente com os achados da fase exploratória, reforçando a robustez da estrutura unidimensional da escala de percepção de aprendizagem. Os índices de ajuste satisfatórios, combinados com cargas fatoriais significativas e estáveis, sustentam o modelo anterior, que propõe um único constructo capaz de representar as dimensões avaliadas. Além disso, os elevados índices de confiabilidade e validade indicam que a escala captura de maneira precisa e consistente os aspectos centrais do fenômeno em análise. Esses resultados confirmam as evidências obtidas na fase exploratória, mas também fornecem uma base empírica sólida para a aplicação da escala em estudos futuros voltados à avaliação de efetividade dos sistemas de gestão de desempenho.

Escala de Percepção de Transferência - AFC

A estrutura unidimensional apresentou resultados de ajuste satisfatórios. Conforme pode ser visto na Tabela 47, os valores de qui-quadrado foram significativos e a razão qui-quadrado por graus de liberdade foi satisfatória. Os índices de CFI, TLI, SRMR e RMSEA suportaram o modelo.

Tabela 47

Índices de ajuste do modelo unifatorial de Transferência

Variável	χ^2 (gl)	χ^2 /gl	CFI	TLI	SRMR	RMSEA (90% IC)
Aprendizagem	177.533* (119)	1,492	0,999	0,999	0,054	0,058 (0,039 – 0,076)

Nota: * $p < 0,001$

A Tabela 48 apresenta as cargas fatoriais dos itens da escala de Transferência.

Tabela 48

Cargas fatoriais de Transferência

Fator	Variável	Estimativa	Desvio Padrão	Valor z	p	95% IC	
						Limite inferior	Limite superior
Transferência	ATRA1	0,765	0,040	19,132	<,001	0,687	0,844
	ATRA2	0,873	0,027	32,498	<,001	0,821	0,926
	ATRA5	0,914	0,016	55,473	<,001	0,881	0,946
	ATRA6	0,792	0,035	22,726	<,001	0,724	0,861
	ATRA7	0,891	0,026	34,709	<,001	0,840	0,941
	ATRA8	0,936	0,018	53,184	<,001	0,901	0,971
	ATRA9	0,720	0,046	15,638	<,001	0,630	0,811
	ATRA10	0,852	0,030	28,441	<,001	0,793	0,910
	ATRA11	0,953	0,012	77,010	<,001	0,930	0,977
	ATRA12	0,802	0,034	23,353	<,001	0,735	0,869
	ATRA13	0,716	0,044	16,064	<,001	0,630	0,802
	ATRA14	0,895	0,022	40,866	<,001	0,852	0,938
	BEME1	0,936	0,015	62,678	<,001	0,907	0,966
	DESE1	0,536	0,061	8,855	<,001	0,417	0,654
	MOTI1	0,879	0,022	39,296	<,001	0,835	0,922
	MOTI3	0,773	0,035	22,195	<,001	0,705	0,841
	MOTI4	0,920	0,017	54,683	<,001	0,887	0,953

Nota: Os valores da estimativa padronizada são os mesmos das estimativas da tabela.

Portanto, os dados coletados demonstraram um excelente ajuste global desse modelo, com altas cargas fatoriais para a maioria dos itens. O modelo reflete adequadamente o construto de transferência da fase exploratória.

Considerando as escalas validadas de maneira exploratória e confirmatória, procedeu-se ao teste de hipóteses no Estudo 2 na seção a seguir.

Resultados do Estudo 2

4.3. Teste de Hipóteses Individuais

A primeira etapa de testes de hipóteses foi realizada pelas correlações das variáveis de percepção individual referentes às hipóteses 1 e 2. A estatística descritiva, correlações e a fidedignidade ω de McDonald são apresentadas na Tabela 49.

Tabela 49

Estatística descritiva, correlações e fidedignidade ω de McDonald na diagonal.

Variável	Média	DP	Reação Afetiva	Reação Cognitiva	Aprendizagem	Transferência
Reação Afetiva	1,526	0,770	0,76			
Reação Cognitiva	2,768	0,869	0,113	0,95		
Aprendizagem	2,567	0,820	-0,048	0,009	0,94	
Transferência	3,350	0,732	0,078	0,519*	-0,004	0,97

Nota: * $p < 0,001$. Fidedignidade ω de McDonald na diagonal da tabela.

Os resultados descritivos sugerem que os respondentes têm alta percepção da Transferência, tendo a maior média ($M = 3,350$; $DP = 0,732$). Os participantes percebem a Reação Afetiva e a Reação Cognitiva ($M = 2,768$; $DP = 0,869$) e a Aprendizagem ($M = 2,567$; $DP = 0,820$). Por fim, a Reação Afetiva é percebida com um valor médio menor, o que era esperado, uma vez que os itens dessa variável em geral são itens reversos ($M = 1,526$; $DP = 0,770$). Todos os construtos tiveram fidedignidade ω de McDonald altamente satisfatória ($\omega > 0,70$).

Os resultados foram, em grande parte, consistentes, porém não foram identificadas correlações entre as medidas de Reação com Aprendizagem (Hipótese 1) e de Aprendizagem com Transferência (Hipótese 2). Assim, ambas as relações foram

excluídas da realização da regressão linear por não apresentarem correlações significativas.

Entretanto, foi identificada uma correlação positiva entre Reação e Transferência (0,283, $p < 0,001$), sobretudo quando analisada a variável Cognitiva (0,519, $p < 0,001$). Portanto, denota-se que quanto mais as reações acerca do sistema de GD fazem sentido e são presentes para o respondente, maior é a sua percepção sobre a transferência.

Sendo assim, essa foi a única relação passível de análise por meio de regressão linear simples, sendo a variável dependente a Transferência e a variável independente a Reação Cognitiva. Esse modelo apresentou uma relação significativa em que a Reação Cognitiva (H_1) explica 26,9% da variância da Transferência ($R^2 = 0,269$; R^2 ajustado = 0,264). Não foram identificadas autocorrelações no modelo, ou seja, os resíduos são independentes. A Tabela 50 apresenta o coeficiente da regressão.

Tabela 50

Regressão simples da variável Reação Cognitiva com Transferência

Modelo	Erro padrão	Coeficiente padronizado (β)	t	p	95% IC Bca	
					Limite inferior	Limite superior
H_0 (intercept)	0,061		55,132	< 0,001	3,230	3,470
H_1 (intercept)	0,174		12,271	< 0,001	1,796	2,486
Reação Cognitiva	0,060	0,519	7,262	< 0,001	0,318	0,556

A Reação Cognitiva explica significativamente a variável Transferência e, com base no coeficiente padronizado, uma percepção elevada sobre aspectos da reação cognitiva está diretamente proporcional a uma percepção de transferência maior ($\beta = 0,519$).

Na Tabela 51 são apresentados os resultados dos testes de hipóteses de forma resumida, com o objetivo de orientar o restante da discussão.

Tabela 51

Resultados do Estudo 2

Hipótese	Resultado	Reação Afetiva	Reação Cognitiva
H1: Reação influencia positivamente na Aprendizagem	Não confirmada	-	-

H2: Aprendizagem influencia positivamente na Transferência.	Não confirmada	-	-
*H3: Reação influencia positivamente na Transferência	Parcialmente suportada	-	x

Nota: *Essa hipótese foi formulada com base nos resultados deste estudo e não havia sido previamente considerada. “Parcialmente suportada” significa que a hipótese foi confirmada, porém somente para um dos fatores. A hipótese é considerada suportada para o fator indicado com “x” na tabela.

A Hipótese 1, que previa uma associação direta entre as reações ao sistema de GD e a aprendizagem, não foi confirmada. Esse resultado contraria a literatura predominante, que sugere que as reações dos funcionários ao sistema influenciam positivamente a aquisição de novas habilidades e atitudes no trabalho (*e.g.* Abdullah & Abraham, 2012; Culbertson et al., 2013; Keeping & Levy, 2000; Van Waeyenberg & Decramer, 2018). No entanto, os dados deste estudo indicam que, para os participantes, as impressões e os sentimentos associadas à GD não são vistos como essenciais para o desenvolvimento de competências.

Embora a literatura destaque a aprendizagem como elemento central que transforma reações ao sistema em resultados concretos para o trabalho (Cislaghi & Zugno, 2019; Keeping & Levy, 2000; Modipane et al., 2019; Schleicher et al., 2019), os achados deste estudo sugerem que a aprendizagem pode não ser percebida pelos participantes como dependente das interações com o sistema de GD.

Além disso, a experiência prévia dos participantes pode ter influenciado a forma como percebem suas competências, atribuindo-as mais a vivências anteriores do que ao próprio sistema. Esse aspecto reforça a ideia de que os sistemas de GD não operam isoladamente, mas interagem com outros elementos do contexto organizacional, como capacitação prévia e suporte gerencial.

Adicionalmente, Schleicher et al. (2019) destacam que, para que um sistema de GD atinja os resultados esperados, há uma grande responsabilidade sobre os gestores, que devem “fazer GD bem”, implicando conhecer e aplicar adequadamente as técnicas de feedback e demonstrar disposição para contribuir para o desenvolvimento de suas equipes. Logo, o sistema de GD pode representar uma oportunidade para os gestores desenvolverem habilidades específicas, como a capacidade de fornecer feedback e esclarecer as atividades para a equipe, que provavelmente não aconteceriam sem que houvesse a obrigatoriedade de um sistema formal de avaliação.

A Hipótese 2, que previa uma relação significativa entre a aprendizagem e a transferência, não foi confirmada neste estudo. Embora a literatura aponte que a

aprendizagem gerada pela GD pode impactar positivamente as atitudes e o desempenho no trabalho, os dados não evidenciaram essa relação. No entanto, esse impacto parece depender de outros fatores e pode não ser percebido de forma direta ou consistente por todos.

Organizações que adotam modelos de GD voltados para o desenvolvimento contínuo, como o uso de feedbacks recorrentes, normalmente reconhecem a importância da aprendizagem como um precursor de mudanças de comportamentos no trabalho. Contudo, essa não é realidade dos modelos de GD utilizados nos órgãos onde esta pesquisa foi realizada, os quais possuem sistemas com características mais burocráticas e proforma.

A única relação significativa encontrada foi a da Hipótese 3, que trata da conexão entre Reação e Transferência. Essa hipótese não foi originalmente proposta neste estudo, mas encontrou suporte parcial, já que a variável Reação Cognitiva se mostrou capaz de explicar a variável Transferência.

Schleicher et al. (2019) já destacavam que as reações afetivas, especialmente as de natureza negativa, embora comuns entre os praticantes, têm pouca ou nenhuma relação com as demais dimensões de efetividade e não explicam a Transferência de maneira significativa. Em contraste, os componentes cognitivos das Reações – como justiça e utilidade percebida – emergem como fatores mais relevantes na relação com Transferência.

Quanto às implicações práticas, mostra-se fundamental não reestruturar um sistema de GD com base nas reações afetivas dos funcionários, uma vez que essas reações não são os principais determinantes da percepção de eficácia do sistema. Murphy (2020) observa que, mesmo quando a maioria dos funcionários expressa insatisfação com o sistema de GD, isso não é suficiente para justificar seu abandono. O autor argumenta que, apesar dos esforços para tornar o sistema justo e útil, os sistemas de GD continuarão a apresentar falhas, sendo, em certa medida, imperfeitos. Schleicher et al. (2019) reforçam que mudanças significativas no sistema de GD devem ser direcionadas pelos aspectos cognitivos, como utilidade, justiça e precisão, que demonstram maior impacto na transferência de conhecimentos e habilidades para o ambiente de trabalho.

Embora as reações também abranjam aspectos positivos e demonstrem uma relação significativa com mudanças comportamentais no ambiente de trabalho, a avaliação da efetividade do sistema de GD não pode se restringir apenas a elas enquanto a única relação significativa. É fundamental considerar que outras variáveis,

especialmente aquelas de natureza contextual, podem influenciar a transferência de conhecimentos e comportamentos, desempenhando um papel preditivo essencial. Assim, avaliar a efetividade da GD exige uma abordagem mais abrangente e integrada.

Apesar de o modelo teórico-integrativo desenvolvido neste estudo ter incorporado múltiplas dimensões e critérios para avaliar a efetividade da GD, as conexões entre as variáveis individuais não apresentaram correlações significativas nas medidas construídas e validadas. Esse resultado, no entanto, não implica que as hipóteses foram definitivamente refutadas. Pelo contrário, é necessário reconhecer que a ausência de relações observáveis pode ser atribuída à influência de variáveis contextuais específicas, que não foram consideradas no modelo atual.

Tanto a literatura de avaliação de treinamento quanto os estudos de GD destacam a importância de fatores contextuais específicos de cada organização – como capacitação do avaliador, estilo de liderança, configuração do sistema de GD, políticas institucionais e até mesmo variáveis econômicas ou políticas externas –, que moldam a efetividade do sistema de GD (Abbad et al., 2000; Decramer et al., 2015; Kuvaas, 2011). Esses fatores, ao não serem incorporados ao modelo, podem interferir nas relações testadas, distorcendo os resultados e obscurecendo a verdadeira dinâmica entre as variáveis individuais.

Portanto, as hipóteses formuladas neste estudo não devem ser entendidas como refutadas. A validade dessas hipóteses pode ser reconsiderada quando variáveis contextuais forem incluídas e testadas de maneira sistemática. Até que isso seja realizado, permanece a incerteza sobre o que realmente interfere na efetividade do sistema de GD. Esse cenário reforça a necessidade de ampliar o modelo integrativo para integrar fatores contextuais, permitindo uma análise mais precisa e representativa da complexidade das relações envolvidas na avaliação da GD.

Por fim, este estudo contribui ao demonstrar que a avaliação da efetividade de GD é multifacetada e que, para avançar na compreensão de seus determinantes, é indispensável expandir a abordagem teórica e metodológica, integrando contextos específicos e variáveis que capturam a realidade organizacional de maneira mais abrangente.

5. CONCLUSÃO

Este estudo teve como objetivo testar empiricamente um modelo de avaliação de efetividade do sistema de Gestão de Desempenho para o contexto da Administração Pública federal brasileira. Para isso, foi proposto um modelo teórico-integrativo, baseado na condução de uma revisão de literatura, que buscou reunir uma série de variáveis de ordem individual e organizacional, comumente adotadas em uma lógica isoladas e sem uma abordagem holística entre elas (Ikramullah et al., 2016; Iqbal et al., 2019; Levy & Williams, 2004a; Schleicher et al., 2019; Zahra & Jadoon, 2016).

Assim, o modelo proposto possibilitou estabelecer relações entre as variáveis, permitindo identificar os fatores que conduzem à efetividade do sistema de GD, além de oferecer uma ferramenta para avaliar se o sistema de GD alcançou os resultados originalmente planejados.

Uma vez que a literatura apresenta medidas de avaliação de efetividade de GD, em geral, extremamente genéricas, impedindo conclusões claras acerca da efetividade de GD, o desenvolvimento e a apresentação de evidências de validade das Escala de Percepção de Reação da Gestão de Desempenho (EPR-GD), Escala de Percepção de Aprendizagem da Gestão de Desempenho (EPA-GD) e Escala de Percepção de Transferência da Gestão de Desempenho (EPT-GD) mostraram-se de grande relevância para o campo, além de ter viabilizado o teste empírico das relações entre essas variáveis do modelo.

Dessa forma, os instrumentos foram construídos inicialmente a partir da tradução de itens de escalas revisadas que deram origem a uma versão inicial dos instrumentos. Essas versões subsidiaram a realização de 11 entrevistas com 14 praticantes de GD em organizações públicas para aprimoramento dos instrumentos de pesquisa inicialmente construídos. Em seguida, foi feita a validação das escalas com nove juízes especialistas no tema.

Essa versão dos instrumentos foi então submetida à validação, inicialmente por meio da Análise Fatorial Exploratória (AFE), conduzida com 238 servidores públicos de organizações de diferentes poderes e esferas, priorizando uma amostra heterogênea. Como resultado, as Escalas de Percepção de Reação, Escalas de Percepção de Aprendizagem e Escala de Transferência apresentaram ajustes adequados e confiáveis, sendo a medida de Reação bidimensional com modelo de segunda-ordem, a qual proporcionou uma visão abrangente, destacando a complementaridade entre Reações

Afetivas e Cognitivas. Já as escalas de Aprendizagem e de Transferência se apresentaram como unidimensionais, divergindo da teoria em que eram previstos múltiplos fatores.

Apesar da divergência, a passagem de um modelo teórico multifatorial para um modelo unifatorial foi suportada pelas evidências empíricas que demonstram itens da escala e representa uma convergência coerente entre fundamentação teórica e evidências estatísticas, proporcionando uma abordagem mais parcimoniosa do instrumento.

Por fim, os instrumentos de pesquisa das três escalas validadas de forma exploratória foram aplicados em dois órgãos públicos, sendo um do Poder Executivo e outra do Poder Judiciário. Os dois órgãos possuem sistemas formais de GD vigentes com características semelhantes: ambos adotam GD para progressão e promoção de carreiras e identificação de necessidades de treinamento e desenvolvimento.

Os dados coletados foram submetidos à Análise Fatorial Confirmatória (AFC) e posteriormente realizado o teste de hipóteses com as respostas dos 145 servidores vinculados a duas organizações públicas federais e que afirmaram já ter participado pelo menos uma vez do processo de GD em sua organização.

Portanto, diferentemente do previsto pela literatura de avaliação da efetividade de GD, as relações entre Reação e Aprendizagem e Aprendizagem e Transferência não foram confirmadas. Em contraste, a análise identificou uma relação significativa entre Reação e Transferência, a qual não havia sido proposta inicialmente no modelo teórico. Essa relação mostra-se relevante principalmente quando considerada as percepções de justiça, precisão e utilidade do sistema, relacionadas às reações cognitivas do modelo, ao contrário das dimensões afetivas, que mostraram impacto limitado.

Os achados deste estudo têm implicações práticas relevantes para gestores públicos e formuladores de políticas de gestão. Apesar das reações cognitivas não serem por si só suficientes para garantir a transferência de comportamentos mais ajustados ao trabalho, destaca-se a necessidade de priorizar a promoção de reações cognitivas – como percepções de justiça, precisão e utilidade – no desenho e implementação dos sistemas de GD. Ficou evidente que focar em sistemas de GD justos, precisos e úteis tem influência positiva na transferência, o que pode implicar que os sistemas também despertem sentimentos menos negativos.

Fato é que, como evidenciado, as reações afetivas, sobretudo aquelas negativas, estarão presentes, não sendo esta razão suficiente para alterar sistemas de GD (Schleicher et al., 2019). Sendo assim, em vez de focar exclusivamente na mitigação de reações afetivas negativas, destaca-se a importância de os gestores investirem em aspectos

contextuais, como o treinamento de avaliadores para a aplicação de feedback de forma justa e construtiva. Essas medidas podem aumentar a eficácia do sistema e garantir maior alinhamento com os objetivos institucionais.

Logo, a ausência de relação direta entre as variáveis Reação, Aprendizagem e Transferência resta evidente para a literatura de gestão de desempenho, algo que há algum tempo já estava claro em outros campos de estudo de gestão de pessoas, como de treinamento (*e.g.* Abbad et al., 2000; Alliger & Janak, 1989; Abbad et al., 2001).

Assim como apontado pela literatura de avaliação de treinamento, este estudo demonstra que o nexos entre essas variáveis provavelmente sofre interferência de outras variáveis, como de natureza contextual, por exemplo (*e.g.* Araújo, 2018; Camões & Meneses, 2016; Côrtes & Meneses, 2019; Araújo et al., 2022). A inclusão de variáveis contextuais no modelo de avaliação da efetividade de GD e exploração de relações mais complexas e não lineares permitirão uma análise mais precisa das interações entre o sistema e o ambiente organizacional, ampliando a compreensão sobre os fatores que influenciam os resultados dos sistemas de GD.

Portanto, enquanto não forem incluídas, no modelo integrativo proposto, as variáveis que operacionalizem aspectos contextuais como interferentes ou estranhos, e enquanto não forem realizados estudos que desenvolvam e testem hipóteses alternativas em diferentes contextos, as suposições formuladas permanecerão sem suporte ou rejeição definitiva. Um exemplo seria o uso de instrumentos específicos, como a Escala de Reputação das Unidades de RH no Setor Público (*e.g.* Escala de Reputação das Unidades de RH no Setor Público, de Souza & Cavalcante, 2023), para testar variáveis em casos específicos.

É fundamental investigar a operacionalização de aspectos contextuais como variáveis interferentes ou estranhas, desenvolvendo formas de traduzi-los em indicadores mensuráveis. Essas relações devem ser analisadas à luz de contextos específicos, permitindo testar diferentes variáveis contextuais e suas interações com o modelo proposto. A condução de testes empíricos em diversos cenários organizacionais, levando em consideração variações no grau de maturidade das práticas de GD entre as organizações públicas, possibilitará uma análise mais aprofundada das condições que favorecem ou limitam a efetividade do sistema.

Destaca-se também que os estudos futuros podem se beneficiar do desenvolvimento e validação dos instrumentos específicos das escalas apresentadas neste estudo. Esses instrumentos poderão ser utilizados tanto para avaliar a efetividade do

sistema de GD sob a perspectiva individual quanto para testar hipóteses em casos específicos, proporcionando suporte empírico adicional ao modelo e permitindo ajustes conforme as características contextuais.

Considerando a recente formulação do constructo de aprendizagem no campo da efetividade dos sistemas de GD, recomenda-se que pesquisas futuras busquem aprimorar essa escala, refinando seus itens, especialmente em relação a aspectos emergentes que possam ter sido sub-representados. Além disso, a aplicação da escala em ambientes diversos permitirá validar sua robustez e explorar como as percepções de aprendizagem variam em função das práticas de gestão de desempenho, que vêm sendo adaptadas para atender a demandas contemporâneas e contextos institucionais dinâmicos.

Com relação às limitações deste estudo, destaca-se que, embora o modelo proposto integre variáveis de ordem individual e organizacional, a ausência de medidas consistentes na literatura limitou o teste empírico a variáveis de natureza individual, como Reação, Aprendizagem e Transferência. Essa limitação impediu a inclusão, desenvolvimento e o teste empírico de variáveis contextuais e organizacionais, ressaltando a necessidade de futuras investigações que ampliem a abrangência do modelo. Assim, reforça-se a importância de continuar testando as hipóteses do modelo, integrando variáveis individuais com variáveis organizacionais, o que permitirá uma análise mais abrangente e contextualizada. Para isso, é essencial o desenvolvimento de escalas que utilizem também medidas objetivas, complementando as abordagens perceptuais e ampliando a capacidade de avaliação dos sistemas de GD.

Uma outra limitação importante deste estudo foi a realização da coleta de dados para o teste de hipóteses em um único momento, o que inviabilizou estabelecer comparações temporais. Uma possibilidade seria a inclusão de estudos longitudinais para compreender a evolução da efetividade dos sistemas de GD a partir de mudanças e alterações empreendidas ao longo do tempo. Outra possibilidade seria aplicação de avaliações em momentos distintos, utilizando pré-testes e pós-testes para capturar os efeitos de mudanças implementadas nos sistemas de GD ao longo do tempo. Dessa forma também poderia ter sido fornecida uma outra abordagem, conforme sugerido por Decramer et al. (2015).

Outro aspecto limitante refere-se às dificuldades metodológicas enfrentadas, especialmente na obtenção de uma quantidade adequada de respostas. Esse desafio impactou tanto a fase de análise exploratória quanto a confirmatória e limitou o tempo disponível para a análise dos dados. A coleta por acessibilidade, viabilizada por meio de

apoio institucional em órgãos nos quais a pesquisadora possuía abertura, restringiu a diversidade da amostra, refletindo, inclusive, na disponibilidade de respostas, uma vez que uma das organizações que apoiou a pesquisa possuía um número reduzido de servidores. Para superar essas limitações em estudos futuros, seria benéfico ampliar a coleta de dados em diferentes órgãos públicos, abrangendo sistemas de GD com características específicas. Isso permitiria obter outras reflexões sobre as relações entre as variáveis estudadas, contribuindo para uma análise mais robusta do impacto dos sistemas de GD em diferentes contextos organizacionais.

Neste estudo, não foi possível avaliar se as diferentes características da amostra influenciam a forma como as relações entre as variáveis são percebidas. Uma investigação futura poderia explorar se gestores e servidores têm percepções distintas sobre a efetividade do sistema de GD, conforme sugerido por Schleicher et al. (2019), que destacaram variáveis específicas capazes de influenciar esses dois grupos de maneira diferenciada.

Adicionalmente, destaca-se a necessidade do estudo da variável “aprendizagem” na ótica dos gestores, pois parece ter um peso maior para eles, considerando que desempenham um papel central na eficácia da GD. Segundo os autores Schleicher et al. (2019), o gestor é fundamental para o sucesso do sistema, sendo responsável por adquirir os conhecimentos necessários para realizar avaliações justas e consistentes, além de aplicar feedback de qualidade.

Outro aspecto relevante para estudos futuros seria a análise dos diferentes desenhos de gestão de desempenho utilizados em organizações públicas. Sabe-se da limitação da aplicação de GD enquanto ferramenta estruturante de GEP nos órgãos, entretanto, seria válido testar o modelo a partir da coleta de dados em uma realidade que incluía não apenas AD, mas também GD. É possível que ao ir além das avaliações tradicionais aplicadas a servidores para progressão e promoção na carreira, as hipóteses testadas teriam mais chances de serem confirmadas, uma vez que os mecanismos de GD refletem dinâmicas e objetivos distintos que podem influenciar a percepção de efetividade do sistema.

Por fim, esta pesquisa contribuiu tanto para o avanço teórico quanto prático do campo da gestão de desempenho ao sistematizar a produção acadêmica sobre avaliação de efetividade de GD em um modelo integrativo. Compreender as relações entre as dimensões e demonstrar empiricamente como a GD se conecta ao desempenho organizacional mostra-se uma ferramenta útil para determinar sua efetividade. Estes

resultados permitem orientar a formulação de sistemas de GD mais eficazes, capazes de gerar benefícios mensuráveis tanto para os indivíduos quanto para as organizações, contribuindo, assim, para superar confusões conceituais e práticas genéricas que limitam o avanço do campo.

REFERÊNCIAS

- Abbad, G. da S., Mourão, L., Meneses, P. P. M., Zerbini, T., Borges-Andrade, J. E., & Vilas-Boas, R. (2012). *Medidas de Avaliação em Treinamento, Desenvolvimento e Educação: Ferramentas para gestão de pessoas*. Artmed.
- Abbad, G., Gama, A. L. G., & Borges-Andrade, J. E. (2000). Treinamento: análise do relacionamento da avaliação nos níveis de reação, aprendizagem e impacto no trabalho. *Revista de Administração Contemporânea*, 4(3), 25–45.
<https://doi.org/10.1590/s1415-65552000000300003>
- Abdullah, H., & Abraham, K. G. (2012). Perception on the performance appraisal system among Malaysian diplomatic officers. *Social Sciences (Pakistan)*, 7(3), 486–495. <https://doi.org/10.3923/sscience.2012.486.495>
- Abu-Shanab, R., Al-Hattami, A., Salman, N., & Ahmed, A. (2021). The Effectiveness of the Performance Appraisal of Public Sector Employees during Covid-19 Pandemic. *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2021*, 685–692.
<https://doi.org/10.1109/3ICT53449.2021.9582025>
- Aguinis, H. (2013). *Performance Management* (3rd ed.). Pearson.
- Ahmed, S. (1999). The emerging measure of effectiveness for human resource management: An exploratory study with performance appraisal. *Journal of Management Development*, 18(6), 543–556.
<https://doi.org/10.1108/02621719910279644>
- Akhtar, Mohammad, & Sushil, S. (2018). Strategic performance management system in uncertain business environment: An empirical study of the Indian oil industry. *Business Process Management Journal*, 24(4), 923–942.
<https://doi.org/10.1108/BPMJ-05-2017-0102>
- Akhtar, Mohd, & Mittal, R. K. (2010). Enterprise performance management and flexibility: Factors of effectiveness in upstream oil industry in india. *Global Journal of Flexible Systems Management*, 11(4), 1–10.
<https://doi.org/10.1007/BF03396589>
- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's Levels of Training Criteria: Thirty Years Later. *Personnel Psychology*, 42(2), 331–342.
<https://doi.org/10.1111/j.1744-6570.1989.tb00661.x>
- Antonialli, F., Antonialli, L. M., & Antonialli, R. (2016). Usos e abusos da escala likert: estudo bibliométrico nos anais do enanpad de 2010 a 2015. *Congresso de*

- Administração, Sociedade e Inovação*, 4494–4516.
- Araújo, L. G. B. de. (2018). *Gestão de Desempenho no Executivo Federal: Iniciativas de enfrentamento aos fatores intervenientes na implementação*. Universidade de Brasília.
- Armstrong, M. (2015). *Armstrong's handbook of performance management: An evidence-based guide to delivering high performance* (2nd ed.).
- Asparouhov, T., & Muthén, B. (2010). Simple second order chi-square correction. *Mplus Technical Appendix*, 1–8.
www.statmodel.com/download/WLSMV_new_chi21.pdf
- Austin, J. T., & Villanova, P. (1992). The Criterion Problem: 1917-1992. *Journal of Applied Psychology*, 77(6), 836–874. <https://doi.org/10.1037/0021-9010.77.6.836>
- Awan, S. H., Habib, N., Shoaib Akhtar, C., & Naveed, S. (2020). Effectiveness of Performance Management System for Employee Performance Through Engagement. *SAGE Open*, 10(4), 1–15.
<https://doi.org/10.1177/2158244020969383>
- Baird, K. (2017). The effectiveness of strategic performance measurement systems. *International Journal of Productivity and Performance Management*, 66(1), 3–21.
<https://doi.org/10.1108/IJPPM-06-2014-0086>
- Baird, K., Su, S. X., & Nuhu, N. (2022). The mediating role of fairness on the effectiveness of strategic performance measurement systems. *Personnel Review*, 51(5), 1491–1517. <https://doi.org/10.1108/PR-07-2020-0573>
- Banks, G. C., & Kepes, S. (2015). The influence of internal HRM activity fit on the dynamics within the “black box.” *Human Resource Management Review*, 25(4), 352–367. <https://doi.org/10.1016/j.hrmr.2015.02.002>
- Bardin, L. (2016). Análise de Conteúdo. São Paulo: Edições 70. In *São Paulo: Edições 70* (p. 279).
- Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17(1), 99–120. <https://doi.org/10.1177/014920639101700108>
- Barney, J. B., & Wright, P. M. (1998). On becoming a strategic partner: The role of human resources in gaining competitive advantage. *Human Resource Management*, 37(1), 31–46. [https://doi.org/10.1002/\(SICI\)1099-050X\(199821\)37:1<31::AID-HRM4>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1099-050X(199821)37:1<31::AID-HRM4>3.0.CO;2-W)
- Barzelay, M., & Jacobsen, A. S. (2009). Theorizing Implementation of Public Management Policy Reforms: A Case Study of Strategic Planning and

- Programming in the European Commission. *Governance*, 22(2), 319–334.
<https://doi.org/10.1111/j.1468-0491.2009.01437.x>
- Becker, B. E., & Huselid, M. A. (2006). Strategic Human Resources Management: Where Do We Go From Here? *Journal of Management*, 32(6), 898–925.
<https://doi.org/10.1177/0149206306293668>
- Bell, V. A., Brady, P. R., Byrne, K. M., Cookson, P. G., Lewis, D. M., & Pailthorpe, M. T. (1983). Investigation Into New Pretreatments for the Printing of Wool and Wool-Blend Fabrics. *Book of Papers, National Technical Conference - AATCC (American Association of Textile Chemists and Colorists)*, 16, 289–298.
- Biulchi, A. F., & Pauli, J. (2012). Avaliação de Desempenho no Serviço Público: A Experiência do Instituto Nacional do Seguro Social na Implantação da Gratificação de Avaliação de Desempenho do Seguro Social - GDASS. *Revista de Administração IMED*, 2(2), 129–137. <https://doi.org/10.18256/2237-7956/raimed.v2n2p129-137>
- Boon, C., Jiang, K., & Eckardt, R. (2024). The Role of Time in Strategic Human Resource Management Research: A Review and Research Agenda. *Journal of Management*, XX(X), 1–40. <https://doi.org/10.1177/01492063241264250>
- Borges-Andrade, J. E., & Abbad, G. S. (1996). Treinamento e desenvolvimento: reflexões sobre suas pesquisas científicas. In *Revista de Administração* (Vol. 31, Issue 2, pp. 112–125).
- Boroujerdi, F. K., Baradaran, S., & Habibi, M. (2016). Identifying and ranking the requirements of employee performance appraisal effectiveness via a comparison of Q methodology results with those of Friedman Method case study: SAIPA company. *International Business Management*, 10, 2172–2217.
- Borsa, J. C., & Seize, M. de M. (2017). Construção e adaptação de instrumentos psicológicos: dois caminhos possíveis. In B. F. D. & J. C. B. (Eds.) (Ed.), *Manual de desenvolvimento de instrumentos psicológicos* (pp. 15–38).
- Boselie, P., Paauwe, J., & Veld, M. (2019). Human resource management and the resource-based view. In Adrian Wilkinson, N. Bacon, S. Snell, & D. Lepak (Eds.), *The SAGE Handbook of Human Resource Management* (pp. 472–493).
- Boselie, P., Van Harten, J., & Veld, M. (2019). A human resource management review on public management and public administration research: stop right there...before we go any further.... *Public Management Review*.
<https://doi.org/10.1080/14719037.2019.1695880>

- Boxall, P., & Purcell, J. (2000). Strategic human resource management: where have we come from and where should we be going? *International Journal of Management Reviews*, 2(2), 183–203. <https://doi.org/10.1111/1468-2370.00037>
- Brasil. (2008). *Lei nº 11.784, de 22 de setembro de 2008*. Dispõe sobre a reestruturação de cargos e carreiras no serviço público federal e institui a Gratificação de Desempenho. Diário Oficial da União, Brasília, DF. Recuperado de http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/111784.htm
- Brasil. (2010). *Decreto nº 7.133, de 19 de março de 2010*. Regulamenta dispositivos da Lei nº 8.112, de 11 de dezembro de 1990, e dispõe sobre critérios e procedimentos para a avaliação de desempenho dos servidores públicos. Diário Oficial da União, Brasília, DF. Recuperado de http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/decreto/d7133.htm
- Brasil. (2020). *Proposta de Emenda Constitucional nº 32, de 2020*. Altera disposições sobre servidores, empregados públicos e organização administrativa. Brasília, DF. Recuperado de https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2262083&fichaAmigavel=nao&utm_source
- Brown, T. (2015). Confirmatory factor analysis for applied research. In *The American Statistician* (2nd ed.). Guilford Publications. <https://doi.org/10.1198/tas.2008.s98>
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). Guilford Publications.
- 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>
- Budd, J. W. (2020). The psychologisation of employment relations, alternative models of the employment relationship, and the OB turn. *Human Resource Management Journal*, 30(1), 73–83. <https://doi.org/10.1111/1748-8583.12274>
- Burke, R. J., & Cooper, C. L. (2007). The Human Resources Revolution: Why Putting People First Matters. In *Leadership & Organization Development Journal* (Vol. 28, Issue 5). <https://doi.org/10.1108/01437730710761788>
- Camões, M. R. de S., & Meneses, P. P. M. (2016). *Gestão de Pessoas no Governo Federal: Análise da Implementação da Política Nacional de Desenvolvimento de Pessoal* (Cadernos,). ENAP.

- Cappelli, P., & Conyon, M. J. (2017). WHAT DO PERFORMANCE APPRAISALS DO? *ILR Review*, 71(1), 1–29. <https://doi.org/10.1177/0019793917698649>
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and Validity Assessment*. SAGE Publications.
- Cascio, W. F. (2015a). Strategic HRM: Too Important for an Insular Approach. *Human Resource Management*, 54(3), 423–426. <https://doi.org/10.1002/hrm.21735>
- Cascio, W. F. (2015b). Strategic HRM: Too Important for an Insular Approach. *Human Resource Management*, 54(3), 423–426. <https://doi.org/10.1002/hrm.21735>
- 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 (Ed.), *Instrumentação psicológica: Fundamentos e práticas* (pp. 506–520). Artmed.
- Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, 83(4), 615–633. <https://doi.org/10.1037/0021-9010.83.4.615>
- Cislaghi, T. P., & Zugno, D. A. (2019). Análise do processo de avaliação de desempenho: Um estudo de caso em uma indústria moveleira do RS. *Revista de Empreendedorismo e Inovação Sustentáveis*, 4(2), 3–22.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319.
- Coelho, F. A., Mourao, L., De Lima, H. K. B., De Velasco, S. M. V., Faiad, C., & Gomes, L. R. (2016). Performance management in the Brazilian federal public administration: How can the logical framework contributes to its effectiveness? *International Business Management*, 10(9), 1658–1667. <https://doi.org/10.3923/ibm.2016.1658.1667>
- Colakoglu, S., Lepak, D. P., & Hong, Y. (2006). Measuring HRM effectiveness: Considering multiple stakeholders in a global context. *Human Resource Management Review*, 16(2), 209–218. <https://doi.org/10.1016/j.hrmr.2006.03.003>
- Connolly, T., Conlon, E. J., & Deutsch, S. J. (1980). Organizational Effectiveness: A Multiple-Constituency Approach. *Academy of Management Review*, 5(2), 211–218. <https://doi.org/10.5465/amr.1980.4288727>
- Cooke, F. L., Xiao, M., & Chen, Y. (2021). Still in search of strategic human resource management? A review and suggestions for future research with China as an

- example. *Human Resource Management*, 60(1), 89–118.
<https://doi.org/10.1002/hrm.22029>
- Cooper, D. R., & Schindler, P. S. (2014). *Business research methods*. McGraw-Hill.
- Côrtes, F. G., & Meneses, P. P. M. (2019). Gestão estratégica de pessoas no Legislativo Federal brasileiro: condições para a implementação. *Revista de Administração Pública*, 53(4), 657–686. <https://doi.org/10.1590/0034-761220180061>
- Costigan, R. D., Insinga, R. C., Berman, J. J., Ilter, S. S., Kranas, G., & Kureshov, V. A. (2009). An Examination of the Relationship of a Western Performance-Management Process to Key Workplace Behaviours in Transition Economies. *Canadian Journal of Administrative Sciences / Revue Canadienne Des Sciences de l'Administration*, 22(3), 255–267. <https://doi.org/10.1111/j.1936-4490.2005.tb00370.x>
- Culbertson, S. S., Henning, J. B., & Payne, S. C. (2013). Performance appraisal satisfaction: The role of feedback and goal orientation. *Journal of Personnel Psychology*, 12(4), 189–195. <https://doi.org/10.1027/1866-5888/a000096>
- Da Luz, L. C. S., & Filgueiredo Filho, D. B. (2018). Fatores que explicam a percepção da efetividade da avaliação do estágio probatório em uma instituição pública federal. *Revista Do Serviço Público*, 69(3), 677–706.
<https://doi.org/10.21874/rsp.v69i3.1793>
- Dalmoro, M., & Vieira, K. M. (2014). DILEMAS NA CONSTRUÇÃO DE ESCALAS TIPO LIKERT: O NÚMERO DE ITENS E A DISPOSIÇÃO INFLUENCIAM NOS RESULTADOS? *Revista Gestão Organizacional*, 6(3).
- Damásio, B. F. (2012). Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica*, 11(2), 213–228.
http://pepsic.bvsalud.org/scielo.php?script=sci_abstract&pid=S1677-04712012000200007
- de Araújo, M. L., Menezes, P. P. M., & Demo, G. (2022). Challenges in Implementing Competency-Based Management in the Brazilian Public Sector: An Integrated Model. *Sustainability (Switzerland)*, 14(22). <https://doi.org/10.3390/su142214755>
- Decramer, A., Audenaert, M., Van Waeyenberg, T., Claeys, T., Claes, C., Vandeveld, S., van Loon, J., & Crucke, S. (2015). Does performance management affect nurses' well-being? *Evaluation and Program Planning*, 49, 98–105.
<https://doi.org/10.1016/j.evalprogplan.2014.12.018>
- Delery, J. E., & Doty, H. D. (1996). Modes to Theorizing in Strategic Human Resource

- Management: Tests of Universalistic, Contingency, and Configurational Performance Predictions. *Academy of Management Journal*, 39(4), 802–835.
<https://doi.org/10.2307/256713>
- Delery, J. E., & Roumpi, D. (2017). Strategic human resource management, human capital and competitive advantage: is the field going in circles? In *Human Resource Management Journal* (Vol. 27, Issue 1, pp. 1–21). Blackwell Publishing Ltd. <https://doi.org/10.1111/1748-8583.12137>
- Den Hartog, D. N., Boselie, P., & Paauwe, J. (2004). Performance management: A model and research agenda. *Applied Psychology*, 53(4), 556–569.
<https://doi.org/10.1111/j.1464-0597.2004.00188.x>
- 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.supp>
- DeNisi, A. S., & Pritchard, R. D. (2006). Performance Appraisal, Performance Management and Improving Individual Performance: A Motivational Framework. *Management and Organization Review*, 2(2), 253–277.
<https://doi.org/10.1111/j.1740-8784.2006.00042.x>
- Dewettinck, K. (2008). Employee performance management systems in Belgian organisations: purpose, contextual dependence and effectiveness. *European Journal of International Management*, 2(2), 192–207.
<https://doi.org/10.1504/EJIM.2008.017761>
- Dewettinck, K., & van Dijk, H. (2013). Linking Belgian employee performance management system characteristics with performance management system effectiveness: exploring the mediating role of fairness. *The International Journal of Human Resource Management*, 24(4), 806–825.
<https://doi.org/10.1080/09585192.2012.700169>
- Dhas, J. M. M., Joseph, S., & Krishnan, S. A. (2015). An empirical study on effectiveness of performance appraisal in reliance fresh Ltd, Chennai. *International Journal of Applied Business and Economic Research*, 13(8), 6433–6445.
- Dyer, L., & Reeves, T. (1995). Human resource strategies and firm performance: What do we know and where do we need to go? *The International Journal of Human Resource Management*, 6(3), 656–670.
<https://doi.org/10.1080/095851995000000041>
- Elvira, E. J. S. (2018). ANÁLISE DA EFETIVIDADE DA POLÍTICA DE AVALIAÇÃO

*DE DESEMPENHO INDIVIDUAL NA ADMINISTRAÇÃO FEDERAL
BRASILEIRA.*

- Espejo, M. M. dos S. B., Araújo, W. S., Fernandes, N., & Domingues, F. (2021). A INFLUÊNCIA DA AVALIAÇÃO DE DESEMPENHO NA ROTATIVIDADE DE PESSOAL DE UMA EMPRESA CONCESSIONÁRIA DE ENERGIA ELÉTRICA NO CENTRO-OESTE. *RMC - Revista Mineira de Contabilidade*, 22(2), 51–64.
- Evans, E. M., & McShane, S. L. (1988). Employee perceptions of performance appraisal fairness in two organizations. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, 20(2), 177–191. <https://doi.org/10.1037/h0079926>
- Farndale, E., McDonnell, A., Scholarios, D., & Wilkinson, A. (2020). The psychologisation conversation: An introduction. *Human Resource Management Journal*, 30(1), 32–33. <https://doi.org/10.1111/1748-8583.12276>
- 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, 762–788.
- Field, A. (2020). *Descobrimos a estatística usando o SPSS* (5th ed.). Penso.
- Fontanella, C. B. J. B., Jose, B., Fontanella, B., Luchesi, B. M., Giovana, M., Saidel, B., Ricas, J., Turato, E. R., & Melo, D. G. (2011). Amostragem em pesquisas qualitativas: proposta de procedimentos para constatar saturação teórica. *Caderno de Saúde Pública*, 27(2), 389–394. <http://www.scielo.br/pdf/csp/v27n2/20.pdf>
- Gardênia Abbad, Jairo Eduardo Borges-Andrade Sallorenzo, L. H., Gamas, A. L. G., & Morandine, D. (2001). Projeto instrucional, aprendizagem, satisfação com o treinamento e auto-avaliação de impacto do treinamento no trabalho. *RPOT*, 1(2), 129–161.
- Godard, J. (2014). The psychologisation of employment relations? *Human Resource Management Journal*, 24(1), 1–18. <https://doi.org/10.1111/1748-8583.12030>
- Godard, J. (2020). Psychologisation revisited. *Human Resource Management Journal*, 30(1), 84–85. <https://doi.org/10.1111/1748-8583.12273>
- Govender, M., & Bussin, M. H. R. (2020). Performance management and employee engagement: A South African perspective. *SA Journal of Human Resource Management*, 18, 1–9. <https://doi.org/10.4102/sajhrm.v18i0.1215>
- Guest, D. E. (1987). Human Resource Management and Industrial Relations. *Journal of*

- Management Studies*, 24(5), 503–521. <https://doi.org/10.1111/j.1467-6486.1987.tb00460.x>
- Guest, D. E. (1997). Human resource management and performance: A review and research agenda. *International Journal of Human Resource Management*, 8(3), 263–276. <https://doi.org/10.1080/095851997341630>
- Haines, V. Y., & St-Onge, S. (2012). Performance management effectiveness: practices or context? *International Journal of Human Resource Management*, 23(6), 1158–1175. <https://doi.org/10.1080/09585192.2011.561230>
- Haines, V. Y., St-Onge, S., & Marcoux, A. (2004). Performance management design and effectiveness in quality-driven organizations. *Canadian Journal of Administrative Sciences*, 21(2), 146–161. <https://doi.org/10.1111/j.1936-4490.2004.tb00330.x>
- Haukoos, J. S., & Lewis, R. J. (2005). Advanced statistics: Bootstrapping confidence intervals for statistics with “difficult” distributions. *Academic Emergency Medicine*, 12(4), 360–365. <https://doi.org/10.1197/j.aem.2004.11.018>
- Hondeghem, A., & Dorpe, K. Van. (2013). Performance management systems for senior civil servants: how strong is the managerial public service bargain? *International Review of Administrative Sciences*, 79(1), 9–27. <https://doi.org/10.1177/0020852312467549>
- Hood, C. (1995). The “new public management” in the 1980s: Variations on a theme. *Accounting, Organizations and Society*, 20(2–3), 93–109. [https://doi.org/10.1016/0361-3682\(93\)E0001-W](https://doi.org/10.1016/0361-3682(93)E0001-W)
- Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, 38(3), 635–672. <https://doi.org/10.2307/256741>
- Huselid, Mark A. (1995). The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance. *Academy of Management Journal*, 38(3), 635–672. <https://doi.org/10.2307/256741>
- Ikramullah, M., Van Prooijen, J. W., 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. (2015). Effectiveness of Performance Appraisal: An Integrated Framework. *International Journal of Management Reviews*, 17(4),

- 510–533. <https://doi.org/10.1111/ijmr.12050>
- 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>
- Ishaq, H. M., Iqbal, M. Z., & Zaheer, A. (2009). Effectiveness of performance appraisal: Its outcomes and detriments in Pakistani organizations. *European Journal of Social Sciences*, 10(3), 479–485.
- Ishaq, H. M., Mansor, N. N. A., Khan, F., & Ahmed, F. (2013). To investigate the factors analysis of effectiveness of performance appraisal and big five personality trait through exploratory factor analysis. *World Applied Sciences Journal*, 26(8), 1030–1038. <https://doi.org/10.5829/idosi.wasj.2013.26.08.13547>
- Jackson, S. E., & Schuler, R. S. (1995a). Understanding Human Resource Management in the Context of Organizations and their Environments. *Annual Review of Psychology*, 46(1), 237–264. <https://doi.org/10.1146/annurev.ps.46.020195.001321>
- Jackson, S. E., & Schuler, R. S. (1995b). Understanding Human Resource Management in the Context of Organizations and their Environments. *Annual Review of Psychology*, 46(1), 237–264. <https://doi.org/10.1146/annurev.ps.46.020195.001321>
- Jackson, S. E., Schuler, R. S., & Jiang, K. (2014). An Aspirational Framework for Strategic Human Resource Management. *The Academy of Management Annals*, 8(1), 1–56. <https://doi.org/10.1080/19416520.2014.872335>
- Jiang, K., & Li, P. (2019). Models of strategic human resource management. In A. Wilkinson, N. Bacon, S. Snell, & D. Lepak (Eds.), *The SAGE Handbook of Human Resource Management* (pp. 23–40).
- Jiang, K., & Messersmith, J. (2018a). On the shoulders of giants: a meta-review of strategic human resource management. *International Journal of Human Resource Management*, 29(1), 6–33. <https://doi.org/10.1080/09585192.2017.1384930>
- Jiang, K., & Messersmith, J. (2018b). On the shoulders of giants: a meta-review of strategic human resource management. *International Journal of Human Resource Management*, 29(1), 6–33. <https://doi.org/10.1080/09585192.2017.1384930>
- Jiang, K., Takeuchi, R., & Lepak, D. P. (2013). Where do we go from here? New perspectives on the black box in strategic human resource management research. *Journal of Management Studies*, 50(8), 1448–1480. <https://doi.org/10.1111/joms.12057>
- Kakkar, S., Dash, S., Vohra, N., & Saha, S. (2020). Engaging employees through

- effective performance management: an empirical examination. *Benchmarking: An International Journal*, 27(5), 1843–1860. <https://doi.org/10.1108/BIJ-10-2019-0440>
- Kang, Y. C., & Choi, N. (2019). Understanding the effectiveness of performance management system: Focus on the perception of South Korean police officers. *Policing*, 42(5), 847–862. <https://doi.org/10.1108/PIJPSM-11-2018-0168>
- Kaufman, B. E. (2015a). Evolution of Strategic HRM as Seen Through Two Founding Books: A 30th Anniversary Perspective on Development of the Field. *Human Resource Management*, 54(3), 389–407. <https://doi.org/10.1002/hrm.21720>
- Kaufman, B. E. (2015b). Evolution of Strategic HRM as Seen Through Two Founding Books: A 30th Anniversary Perspective on Development of the Field. *Human Resource Management*, 54(3), 389–407. <https://doi.org/10.1002/hrm.21720>
- Kaufman, B. E. (2015c). The RBV theory foundation of strategic HRM: Critical flaws, problems for research and practice, and an alternative economics paradigm. *Human Resource Management Journal*, 25(4), 516–540. <https://doi.org/10.1111/1748-8583.12085>
- Kaufman, B. E. (2020). The real problem: The deadly combination of psychologisation, scientism, and normative promotionalism takes strategic human resource management down a 30-year dead end. *Human Resource Management Journal*, 30(1), 49–72. <https://doi.org/10.1111/1748-8583.12278>
- Kaufman, R., & Keller, J. M. (1994). Levels of Evaluation : Beyond Kirkpatrick. *Human Resource Development Quarterly*, 5(4), 71–80.
- Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. *Journal of Applied Psychology*, 85(5), 708–723. <https://doi.org/10.1037/0021-9010.85.5.708>
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3).
- Kepes, S., & Delery, J. E. (2006). Designing Effective HRM Systems: The Issue of HRM Strategy. In *The Human Resources Revolution: Why Putting people first matters* (pp. 55–76).
- Kim, J. (2016). Impact of Performance Appraisal Justice on the Effectiveness of Pay-for-Performance Systems After Civil Service Reform. *Public Personnel Management*, 45(2), 148–170. <https://doi.org/10.1177/0091026016644625>
- Kinicki, A. J., Jacobson, K. J. L., Peterson, S. J., & Prussia, G. E. (2013). Development

- and Validation of the Performance Management Behavior Questionnaire. *Personnel Psychology*, 66(1), 1–45. <https://doi.org/10.1111/peps.12013>
- 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>
- Kraiger, K., & Ford, J. K. (2021). The Science of Workplace Instruction: Learning and Development Applied to Work. *Annual Review of Organizational Psychology and Organizational Behavior*, 8, 45–72. <https://doi.org/10.1146/annurev-orgpsych-012420-060109>
- Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation. *Journal of Applied Psychology*, 78(2), 311–328. <https://doi.org/10.1037/0021-9010.78.2.311>
- Kuipers, B. S., & Giurge, L. M. (2016). Does alignment matter? The performance implications of HR roles connected to organizational strategy. *The International Journal of Human Resource Management*, 1–23. <https://doi.org/10.1080/09585192.2016.1155162>
- Kuvaas, B. (2011). The interactive role of performance appraisal reactions and regular feedback. *Journal of Managerial Psychology*, 26(2), 123–137. <https://doi.org/10.1108/02683941111102164>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Lawler, E. E. (2003). Reward Practices and Performance Management System Effectiveness. *Organizational Dynamics*, 32(4), 396–404. <https://doi.org/10.1016/j.orgdyn.2003.08.007>
- Lawler, E. E., & McDermott, M. (2003). Current performance management practices. *WorldatWork Journal*, 12(2), 49.
- Lengnick-Hall, M. L., Lengnick-Hall, C. A., Andrade, L. S., & Drake, B. (2009). Strategic human resource management: The evolution of the field. *Human Resource Management Review*, 19, 64–85. <https://doi.org/10.1016/j.hrmr.2009.01.002>
- Lepak, D. P., Liao, H., Chung, Y., & Harden, E. E. (2006). A Conceptual Review of Human Resource Management Systems in Strategic Human Resource

- Management Research. In *Research in Personnel and Human Resources Management* (Vol. 25, pp. 217–271). [https://doi.org/10.1016/S0742-7301\(06\)25006-0](https://doi.org/10.1016/S0742-7301(06)25006-0)
- Levy, P. E., Tseng, S. T., Rosen, C. C., & Lueke, S. B. (2017). Performance management: A marriage between practice and science - Just say “i do.” In *Research in Personnel and Human Resources Management* (Vol. 35, Issue July, pp. 155–213). Emerald. <https://doi.org/10.1108/S0742-730120170000035005>
- Levy, P. E., & Williams, J. R. (2004a). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30(6), 881–905. <https://doi.org/10.1016/j.jm.2004.06.005>
- Levy, P. E., & Williams, J. R. (2004b). The Social Context of Performance Appraisal: A Review and Framework for the Future. *Journal of Management*, 30(6), 881–905. <https://doi.org/10.1016/j.jm.2004.06.005>
- 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>
- Longenecker, C. O., & Nykodym, N. (1996). Public Sector Performance Appraisal Effectiveness: A Case Study. *Public Personnel Management*, 25(2), 151–164. <https://doi.org/10.1177/009102609602500203>
- Lopes, A. de O. B., Araújo, R. M. de, & Costa, M. T. P. (2020). GESTÃO DE DESEMPENHO INDIVIDUAL NO GOVERNO FEDERAL: um survey com agentes implementadores em uma organização pública. *Revista Interface*, 83–105.
- Lopes, F. J. (2020). COMPETENCY MANAGEMENT , KNOWLEDGE MANAGEMENT AND CORPORATIVE EDUCATION : A STUDY ON BRAZILIAN GESTÃO POR COMPETÊNCIAS , GESTÃO DO CONHECIMENTO E EDUCAÇÃO CORPORATIVA : UM ESTUDO SOBRE EMPRESAS BRASILEIRAS. *Journal on Innovation and Sustainability*, 11, 147–158.
- MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. H. (Ed.) (Ed.), *Structural equation modeling: Concepts, issues, and applications*. Sage.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84–99. <https://doi.org/10.1037/1082-989X.4.1.84>

- Makhubela, M., Botha, P. A., & Swanepoel, S. (2016). Employees' perceptions of the effectiveness and fairness of performance management in a South African public sector institution. *SA Journal of Human Resource Management*, 15, 1–11. <https://doi.org/10.4102/sajhrm.v14i1.728>
- Maley, J. F., & Moeller, M. (2014). Global performance management systems: The role of trust as perceived by country managers. *Journal of Business Research*, 67(1), 2803–2810. <https://doi.org/10.1016/j.jbusres.2012.08.003>
- Martins, B., Marques, J., Vieira, L., & Garay, A. (2009). Avaliação de Desempenho Individual no Setor Público Brasileiro: Análise da Produção Acadêmica de 2000 a 2009. *IV Encontro De Administração Pública e Governança*, 1–17.
- Martins, V. A., & Ensslin, S. R. (2020). Avaliação de Desempenho no Setor Público: Oportunidades de Pesquisa com Base nas Lacunas Identificadas em Estudos de Casos. *REAd. Revista Eletrônica de Administração (Porto Alegre)*, 25(3), 3–22.
- Modipane, P. I., Botha, P. A., & Blom, T. (2019). Employees' perceived effectiveness of the performance management system at a north-west provincial government department. *South African Journal of Childhood Education*, 17. <https://doi.org/10.4102/sajhrm.v17i0.1081>
- 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>
- Monsur, M. P., & Akkas, M. A. (2015). Performance appraisal system of the academics of public and private universities of Bangladesh: An empirical study. *International Business Management*, 9(5), 712–718. <https://doi.org/10.3923/ibm.2015.712.718>
- Motowildo, S. J., Borman, W. C., & Schmit, M. J. (1997). A Theory of Individual Differences in Task and Contextual Performance. *Human Performance*, 10(2), 71–83. https://doi.org/10.1207/s15327043hup1002_1
- Murphy, K. R., & Cleveland, J. N. (1995). Understanding performance appraisal: social, organizational, and goal-based perspectives. In *Choice Reviews Online* (Vol. 32, Issue 10). <https://doi.org/10.5860/choice.32-5748>
- Murphy, Kevin R. (2020). Performance evaluation will not die, but it should. *Human Resource Management Journal*, 30(1), 13–31. <https://doi.org/10.1111/1748-8583.12259>
- Nagel, S. S. (1986). Efficiency, Effectiveness, and Equity in Public Policy Evaluation.

- Review of Policy Research*, 6(1), 99–120. <https://doi.org/10.1111/j.1541-1338.1986.tb00651.x>
- Nankervis, A. R., Stanton, P., & Foley, P. (2012). Exploring the rhetoric and reality of performance management systems and organisational effectiveness - Evidence from Australia. *Research and Practice in Human Resource Management*, 20(1), 40–56.
- Neiva, E. R. (2020). Caderno 6 - Gestão e Avaliação de Desempenho no Setor Público Brasileiro: aspectos críticos e discussão sobre as práticas correntes. *Cadernos Da Reforma Administrativa*, 6.
- Odelius, C. C., & Santos, P. R. G. dos. (2007). Avaliação de desempenho individual na administração pública federal: aspectos intervenientes no processo e nos resultados. *Revista Economia e Gestão*, 7(15), 10–30.
- Ogbonnayaa, C., Daniels, K., Messersmithc, J., & Rofcanin, Y. (2023). A Theory-Based Analysis of Null Causality between HRM Practices and Outcomes: Evidence from Four- Wave Longitudinal Data. *Journal of Management Studies*, 60(6), 1445–1484.
- Ostroff, C., & Bowen, D. E. (2016). Reflections on the 2014 Decade Award: Is There Strength in the Construct of HR System Strength? *Academy of Management Review*, 41(2), 196–214. <https://doi.org/10.5465/amr.2015.0323>
- Paauwe, J. (2011). HRM and performance: Achieving long term viability. *HRM and Performance: Achieving Long Term Viability*, 46(1), 1–274. <https://doi.org/10.1093/acprof:oso/9780199273904.001.0001>
- Paauwe, J., & Boselie, P. (2005). HRM and Performance: What's Next? *CAHRS Working Paper*, 05(09). <http://digitalcommons.ilr.cornell.edu/cahrswp>
- Paauwe, J., & Richardson, R. (1997). Introduction. *The International Journal of Human Resource Management*, 8(3), 257–262. <https://doi.org/10.1080/095851997341621>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Research Methods And Reporting*, 372. <https://doi.org/10.1136/bmj.n71>
- Pasquali, L. (2010). *Teorias e Métodos de Medida em Ciências do Comportamento*.
- Phuong, T. H. (2018). Perceived justice in performance appraisal among Vietnamese

- employees: Antecedents and consequences. *International Journal of Business Excellence*, 15(2), 209–221. <https://doi.org/10.1504/IJBEX.2018.091920>
- Pichler, S. (2012). The social context of performance appraisal and appraisal reactions: A meta-analysis. *Human Resource Management*, 51(5), 709–732. <https://doi.org/10.1002/hrm.21499>
- Pinto, J. de F., & Behr, R. R. (2015). Contradições na avaliação de desempenho dos servidores técnico-administrativos em educação na universidade pública. *Cadernos EBAPE.BR*, 13(4), 795–820. <https://doi.org/10.1590/1679-395132365>
- Ponmuthumari, S. (2020). A study on effectiveness of performance appraisal system in manufacturing sector. *International Journal of Advanced Science and Technology*, 29(2), 3733–3738.
- Latinaud, P., & Marchan, P. (2012). L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française. *11èmes Journées Internationales d'Analyse Statistique Des Données Textuelles*.
- 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>
- Roberts, G. E. (1995). Municipal Government Performance Appraisal System Practices: Is the Whole Less Than the Sum of its Parts? *Public Personnel Management*, 24(2), 197–221. <https://doi.org/10.1177/009102609502400209>
- Roberts, G. E. (1996). A Case Study in Performance Appraisal System Development: Lessons from a Municipal Police Department. *The American Review of Public Administration*, 26(3), 361–385. <https://doi.org/10.1177/027507409602600307>
- Rolle, C., & Klingner, D. (2012). Performance Appraisal Systems as a Strategic Human Resource Management Tool in the Bahamian Public Service. *International Journal of Public Administration*, 35(5), 303–314. <https://doi.org/10.1080/01900692.2012.655459>
- Roumpi, D., & Delery, J. E. (2019). Strategic HRM: Where Do We Go from Here? In Adrian Wilkinson, N. Bacon, S. Snell, & D. Lepak (Eds.), *The SAGE Handbook of Human Resource Management* (pp. 423–438).
- Sadiyawati, M., & Hasanati, N. (2024). Effectiveness of Performance Appraisal System in Employee Development: A Systematic Review. *International Journal of Research and Innovation in Social Science (IJRISS)*, 8(7), 1175–1189. <https://doi.org/10.47772/IJRISS>

- Schleicher, D. J., Baumann, H. M., Sullivan, D. W., & Junhyok, Y. (2019). Evaluating the effectiveness of performance management: A 30-year integrative conceptual review. *Journal of Applied Psychology, 104*(7), 851–887.
<https://doi.org/10.1037/apl0000368>
- 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>
- Schuler, R., & E. Jackson, S. (2014). Human resource management and organizational effectiveness: yesterday and today. *Journal of Organizational Effectiveness, 1*(1), 35–55. <https://doi.org/10.1108/JOEPP-01-2014-0003>
- Secchi, L. (2009). Modelos organizacionais e reformas da administração pública. *Revista de Administração Pública, 43*(2), 347–369.
<http://bibliotecadigital.fgv.br/ojs/index.php/rap/article/viewFile/6691/5274>
- Secchi, L., Farranha, A. C., Rodrigues, K. F., Bergue, S. T., & Medeiros-Costa, C. C. (2021). Reforma Administrativa No Brasil: Passado, Presente E Perspectivas Para O Futuro Frente À Pec 32/2020. *Cadernos Gestão Pública e Cidadania, 26*(83), 1–17. <https://doi.org/10.12660/cgpc.v26n83.82430>
- Selden, S., & Sowa, J. E. (2011). Performance management and appraisal in human service organizations: Management and staff perspectives. In *Public Personnel Management* (Vol. 40, Issue 3, pp. 251–264). SAGE Publications Inc.
<https://doi.org/10.1177/009102601104000305>
- 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/ER-01-2015-0006>
- Sharma, N. P., Sharma, T., & Agarwal, M. N. (2021). Relationship between perceived performance management system (PMS) effectiveness, work engagement and turnover intention: mediation by psychological contract fulfillment. *Benchmarking*.
<https://doi.org/10.1108/BIJ-01-2021-0008>
- Snell, S. A., Youndt, M. A., & Wright, P. M. (1996). Establishing a framework for research in strategic human resource management: Merging resource theory and organizational learning. *Research in personnel and human resources management. Research in Personnel and Human Resources Management, 14*(61–90).

- Soltani, E., Van Der Meer, R., Gennard, J., & Williams, T. (2003). A TQM approach to HR performance evaluation criteria. *European Management Journal*, 21(3), 323–337. [https://doi.org/10.1016/S0263-2373\(03\)00052-5](https://doi.org/10.1016/S0263-2373(03)00052-5)
- Sotirakou, T., & Zeppou, M. (2006). Utilizing performance measurement to modernize the Greek public sector. *Management Decision*, 44(9), 1277–1304. <https://doi.org/10.1108/00251740610707730>
- Souza, I. G. L. de, & Cavalcante, P. P. M. M. (2023). Construção e validação de uma medida de reputação das áreas de gestão de pessoas de organizações públicas brasileiras. *Revista de Administração Pública*, 57(6), 1–19. <https://doi.org/10.1590/0034-761220220329>
- Stanton, P., & Nankervis, A. (2011). Linking strategic HRM, performance management and organizational effectiveness: perceptions of managers in Singapore. *Asia Pacific Business Review*, 17(1), 67–84. <https://doi.org/10.1080/13602381003790382>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Pearson.
- 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>
- Troth, A. C., & Guest, D. E. (2020). The case for psychology in human resource management research. *Human Resource Management Journal*, 30(1), 34–48. <https://doi.org/10.1111/1748-8583.12237>
- Tuytens, M., & Devos, G. (2012). Importance of system and leadership in performance appraisal. *Personnel Review*, 41(6), 756–776. <https://doi.org/10.1108/00483481211263692>
- Ulrich, D. (1998). Measuring human resources: an overview of practice and a prescription for results. *Human Resource Management*, 36(3), 303–320.
- Upadhyay, R. K., Ansari, K. R., & Bijalwan, P. (2020). Performance Appraisal and Team Effectiveness: A Moderated Mediation Model of Employee Retention and Employee Satisfaction. *Vision: The Journal of Business Perspective*, 24(4), 395–405. <https://doi.org/10.1177/0972262919875542>
- Valentini, F., & Damásio, B. F. (2016). Average Variance Extracted and Composite Reliability: Reliability Coefficients. *Psicologia: Teoria e Pesquisa*, 32(2), 1–7. <https://doi.org/10.1590/0102-3772e322225>
- Van Waeyenberg, T., & Decramer, A. (2018). Line managers' AMO to manage

- employees' performance: the route to effective and satisfying performance management. *The International Journal of Human Resource Management*, 29(22), 3093–3114. <https://doi.org/10.1080/09585192.2018.1445656>
- Vilhena, R., & Martins, H. (2022). Percepções sobre gestão do desempenho no setor público. *Revista Do Serviço Público*, 73(Especial Consad), 8–25. <https://doi.org/https://doi.org/10.21874/rsp.v73.ia.6864>
- Wright, A. L., Zammuto, R. F., Liesch, P. W., Middleton, S., Hibbert, P., Burke, J., & Brazil, V. (2016). Evidence-based Management in Practice: Opening up the Decision Process, Decision-maker and Context. *British Journal of Management*, 27(1), 161–178. <https://doi.org/10.1111/1467-8551.12123>
- Wright, P. M., & McMahan, G. C. (1992). Theoretical perspectives for strategic human resource management. In *Journal of management* (Vol. 18, Issue 2, pp. 295–320). Sage PublicationsSage CA: Thousand Oaks, CA. <https://doi.org/10.1177/014920639201800205>
- Wright, P. M., Dunford, B. B., & Snell, S. a. (2001). Human resources and the resource based view of the firm. In *Journal of Management* (Vol. 27, p. 701). [https://doi.org/10.1016/S0149-2063\(01\)00120-9](https://doi.org/10.1016/S0149-2063(01)00120-9)
- Wright, Patrick M., Dunford, B. B., & Snell, S. A. (2001). Human resources and the resource based view of the firm. *Journal of Management*, 27(6), 701–721. <https://doi.org/10.1177/014920630102700607>
- Wright, Patrick M., McMahan, G. C., & McWilliams, A. (1994). Human resources and sustained competitive advantage: A resource-based perspective. *The International Journal of Human Resource Management*, 5(2), 301–326. <https://doi.org/10.1080/095851994000000020>
- Wright, Patrick M., & Ulrich, M. D. (2017). A Road Well Traveled: The Past, Present, and Future Journey of Strategic Human Resource Management. *Annual Review of Organizational Psychology and Organizational Behavior*, 4(1), 45–65. <https://doi.org/10.1146/annurev-orgpsych-032516-113052>
- Wright, R. P., & Cheung, F. K. K. (2007). Articulating appraisal system effectiveness based on managerial cognitions. *Personnel Review*, 36(2), 206–230. <https://doi.org/10.1108/00483480710726118>
- Yang, K., & Hsieh, J. Y. (2007). Managerial Effectiveness of Government Performance. *Public Administration Review*, 67(5), 861–879.
- Yu, Y., Baird, K. M., & Tung, A. (2018). Human resource management in Australian

- hospitals: the role of controls in influencing the effectiveness of performance management systems. *The International Journal of Human Resource Management*, 1–26. <https://doi.org/10.1080/09585192.2018.1511618>
- Zahra, A., & Jadoon, M. Z. I. (2016). Autonomy of public agencies in Pakistan: does structure matter? *International Journal of Public Sector Management*, 29(6), 565–581. <https://doi.org/10.1108/IJPSM-02-2015-0019>
- Zhang, X., Tantardini, M., Kim, Y., & de Lancer Julnes, P. (2020). Is Performance Management Reform Living up to its Expectations? An Analysis of Public Management Reform in China's Guangdong Province. *Public Performance & Management Review*, 1–27. <https://doi.org/10.1080/15309576.2020.1806085>

Apêndice A - Carta de Apresentação da Entrevista (e-mail)

Prezado(a),

O motivo do meu contato hoje é referente à minha pesquisa de doutorado na Universidade de Brasília (UnB), sob orientação do Prof. Dr. Pedro Menezes, que estou realizando no âmbito da Administração Pública Federal.

A pesquisa intitulada “Proposição de um Modelo Integrativo de Avaliação da Efetividade do Sistema de Gestão de Desempenho” visa testar empiricamente um modelo de avaliação da efetividade do sistema de Gestão de Desempenho adaptado ao contexto da Administração Pública Federal brasileira.

Para tanto, estou realizando a coleta de dados, inicialmente, por meio da realização de entrevistas, com duração prevista de 25 a 45 minutos, a depender das respostas fornecidas, com gestores de unidades de GP do setor público que lideram, conhecem e/ou participem do processo de Gestão ou Avaliação de Desempenho na organização.

Assim, uma vez apresentadas as motivações, gostaria de saber se posso contar com a contribuição de vocês nesta etapa da pesquisa. Se sim, solicito-lhes, por gentileza, indicar os dias e horários que melhor se adéquem às atividades de vocês e dos possíveis participantes, de forma que eu possa criar o convite da reunião virtual.

Para eventuais dúvidas ou necessidade de contato, por favor, responder a este endereço de e-mail.

Agradeço desde já pela atenção e estou à disposição.

Atenciosamente,

Mariana Lopes

Apêndice B – Instrumento 1 – Roteiro de Entrevista

ROTEIRO DE ENTREVISTA

Avaliação da Efetividade de Sistemas de Gestão De Desempenho⁵

Prezado(a) entrevistado(a),

Este roteiro é um instrumento de coleta de dados que compõe o projeto de pesquisa de doutorado em Administração da aluna Msc. Mariana Lopes de Araújo, sob a orientação do Prof. Doutor Pedro Paulo Murce Menezes, na Universidade de Brasília (UnB).

Seção 1: Sistema de GD/AD em geral

1. Em geral, quais fatores ou aspectos você considera importantes em um sistema de GD/AD?
2. O que um sistema de GD/AD tem que ter ou ser para ser considerado um sistema efetivo?
3. Na sua opinião, quais resultados ou consequências decorrem de um sistema de GD/AD de sucesso?

Seção 2: Critérios e dimensões

Sabemos que pela literatura, alguns aspectos são importantes: ser justo, satisfatório, útil, melhorar o feedback, promover desempenho superior no trabalho etc.).

1. O que você teria a dizer sobre estes aspectos no sistema de GD/AD da organização?
2. Como as pessoas da organização esperam que o sistema de GD/AD seja? E o que elas dizem sobre o sistema da organização?
3. Quais são suas percepções sobre o sistema de GD/AD da organização? E dos demais servidores?
4. As pessoas normalmente visualizam quais aprendizados, resultados ou consequências decorrentes do sistema de GD/AD?

Seção 3: Classificação dos itens traduzidos para realidade organizacional

Pensando na realidade da sua organização e do contexto em que ela está inserida (Poder Executivo Federal, setor público etc.), discorra sobre a relevância do itens a seguir:

⁵ Gestão de Desempenho (GD) e Avaliação de Desempenho (AD) serão tratados de forma intercambiável neste estudo.

Dimensão Reação: Impressões, julgamentos, avaliações e opiniões em geral dos colaboradores e gestores sobre o sistema de GD após a participação ou contato com ações ou práticas relacionadas ao sistema de GD.

Nº	Critério	Definição	Itens
1	Justiça	Quanto os colaboradores e os gestores percebem que os resultados obtidos a partir do sistema de GD são consistentes com as normas, garantindo equidade e igualdade; o processo adere a critérios de consistência, ausência de vieses, representação, precisão e ética e permite um espaço participativo na tomada de decisão ou na influência sobre o resultado; a relação entre o avaliado e o avaliador é respeitosa e sensível, com clareza nos processos de tomada de decisão.	A avaliação que recebi do meu desempenho foi condizente com o que eu merecia
2			A avaliação que recebi do meu desempenho aferiu o que eu tenho contribuído para a organização
3			A avaliação que recebi do meu desempenho forneceu uma avaliação adequada do trabalho que realizei
4			A avaliação que recebi do meu desempenho me forneceu uma boa avaliação do esforço que eu tenho colocado no meu trabalho.
5			A avaliação que recebi do meu desempenho foi justa
6			Eu concordo com a avaliação final que eu recebi do meu desempenho
7			A avaliação que recebi é justificada, dado o meu desempenho
8			Os benefícios, recompensas e gratificações recebidos são condizentes com o meu desempenho
9			Sou capaz de expressar minhas opiniões e sentimentos sobre os procedimentos de avaliação de desempenho da minha organização
10			Tenho influência sobre as avaliações de desempenho feitas como resultado dos procedimentos da minha organização
11			Os procedimentos de avaliação de desempenho usados em minha organização foram aplicados consistentemente
12			Os procedimentos de avaliação de desempenho usados em minha organização são livres de preconceito
13			Os procedimentos de avaliação de desempenho usados em minha organização são baseados em informações precisas
14			Posso apelar das avaliações feitas pelos procedimentos usados em minha organização.
15			Os procedimentos de avaliação de desempenho utilizados na minha organização respeitam os padrões éticos e morais
16			O processo e os procedimentos de avaliação do meu desempenho são justos
17			Minha avaliação de desempenho foi conduzida de forma justa
18			De forma geral, a forma com que o processo de avaliação de desempenho foi feito é justo
19			Eu compreendo as razões da minha avaliação de desempenho mais recente
20			Os padrões usados para avaliar meu desempenho são apropriados
21			Eu participo do estabelecimento de padrões e metas usados para avaliar meu desempenho no trabalho
22			Entendo o que devo fazer para receber uma avaliação de desempenho elevada
23			Tenho oportunidades suficientes (como tarefas ou projetos desafiadores) para obter uma avaliação de desempenho elevada
24			Eu sei como minha classificação de desempenho se compara a outras pessoas em minha organização que têm trabalhos semelhantes
25			Medidas objetivas são usadas para avaliar meu desempenho
26			Confio no meu supervisor para avaliar de forma justa o meu desempenho e as minhas contribuições
27			Eu confio que meu supervisor comunique claramente as expectativas de conduta
28			O meu supervisor mantém-me informado sobre o meu desempenho
29			Meu supervisor fornece feedback construtivo sobre meu desempenho no trabalho
30			Meu supervisor fornece feedback oportuno sobre meu desempenho no trabalho
31			Meu supervisor oferece orientação, oportunidades de treinamento ou outra assistência para me ajudar a melhorar minhas habilidades e desempenho
32			Entendo como meu supervisor avaliará meu desempenho.
33			Meu supervisor avalia meu desempenho de forma justa e precisa
34			Meu supervisor é responsável por avaliar o desempenho dos funcionários de forma justa e precisa

35			Meu supervisor me trata de maneira educada
36			Meu supervisor me trata com dignidade
37			Meu supervisor se abstém de observações ou comentários impróprios
38	Precisão	Quanto os colaboradores e gestores percebem que o sistema de GD é preciso, confiável em relação ao planejamento, alinhando metas individuais e organizacionais; feedback e o aconselhamento contínuos; desempenho mensurado e alinhado às expectativas; e aos resultados decorrentes do sistema de GD (gratificações, benefícios, capacitação etc.).	O planejamento do desempenho baseado no sistema de GD dá uma ideia clara do que se espera de mim para atingir as metas organizacionais
39			O plano de desempenho me ajuda a concentrar meus esforços por meio da identificação de metas (e/ou comportamentos/habilidades) relevantes para atingir as metas organizacionais
40			Meu gerente e eu atualizamos minhas metas conforme as metas da organização mudam
41			O feedback contínuo durante o ciclo de desempenho fornece uma avaliação precisa de como estou me saindo em relação ao desempenho planejado
42			Durante o ano, minhas áreas de melhoria são claramente apontadas para mim
43			Recebo o treinamento necessário durante o ano para atingir meus objetivos (e/ ou melhorar meus comportamentos/ habilidades) para atingir o desempenho planejado
44			O feedback anual durante a avaliação de desempenho é uma representação precisa do feedback contínuo durante o ciclo de desempenho
45			Meus objetivos (comportamentos/ habilidades) são avaliados com precisão como parte do processo de revisão
46			Minha avaliação anual de desempenho é muito objetiva na avaliação do meu desempenho anual em relação ao desempenho planejado
47			A avaliação de desempenho resulta em uma classificação de desempenho precisa
48			Os resultados da minha avaliação de desempenho (ou sistema de GD) (compensação, recompensa e/ou reconhecimento) estão vinculados à minha classificação de desempenho
49			Minha avaliação anual de desempenho está diretamente relacionada aos resultados da minha avaliação de desempenho (compensação, recompensa e/ou reconhecimento)
50			As pontuações que recebi foram uma avaliação precisa do meu desempenho
51			As avaliações dos meus avaliadores refletem o meu verdadeiro desempenho
52			A avaliação que recebi do meu desempenho pelos meus avaliadores correspondeu à minha própria avaliação
53	Satisfação	Quanto os colaboradores e gestores se sentem contentes e confortáveis, ocasionando uma reação positiva em geral em relação ao sistema de GD.	Estou satisfeito com a forma como a minha organização me fornece feedback
54			Minha organização é boa em fornecer reconhecimento por bom desempenho
55			O feedback que recebo concorda com o que realmente alcancei
56			Acho que minha organização tenta conduzir a avaliação de desempenho da melhor maneira possível
57			Minha organização parece mais empenhada em fornecer feedback positivo para um bom desempenho do que criticar o mau desempenho
58			A avaliação de desempenho é valiosa para mim e para minha organização
59			As reuniões regulares de avaliação de desempenho que tenho com meu gerente incluem uma discussão sobre as coisas que meu gerente poderia fazer para me ajudar a ter um desempenho melhor
60			O atual sistema de avaliação de desempenho é justo e imparcial
61			Minha última revisão anual de avaliação de desempenho conflitava significativamente com o feedback que recebi em minhas revisões regulares ao longo do ano
62			Minha visão do meu desempenho foi levada em consideração pelo meu gerente ao avaliar o resultado da minha avaliação de desempenho no ano passado

63			O resultado que recebo em minha avaliação anual de desempenho determina meu salário em minha organização
64			Entendo, por meio de minha análise de avaliação de desempenho, como meu trabalho ajuda minha organização a atingir seus objetivos estratégicos
65			Como parte do sistema de avaliação de desempenho da minha organização, recebo feedback regular sobre meu desempenho
66			Meu gerente está totalmente comprometido com minhas avaliações de desempenho
67			O sistema geral de avaliação de desempenho me ajuda a identificar áreas para melhorar meu desempenho no trabalho
68			O resultado que recebo em minha avaliação anual de desempenho determina meu bônus/incentivo
69			Eu decidi sobre as metas que eu tinha que alcançar como parte de minha última avaliação de desempenho, em consulta com meu gerente
70			Minha revisão de avaliação de desempenho de final de ano é a única vez que recebo feedback sobre meu desempenho
71			Todos os envolvidos no sistema de avaliação de desempenho devem receber treinamento sobre como fazê-lo, mesmo que não sejam gerentes
72			Se eu não concordar com o resultado da minha avaliação de desempenho, existe um processo de apelação claro para eu usar
73			No geral, estou satisfeito com a maneira como meu gerente usa o sistema de avaliação de desempenho
74			No geral, estou satisfeito com o sistema de avaliação
75			Quando faço um bom trabalho, recebo feedback positivo do meu gerente
76			Acredito que as metas que tive que alcançar como parte de minha última avaliação de desempenho foram justas e alcançáveis
77			O sistema de avaliação de desempenho que minha organização usa permite a entrada de outras fontes, como colegas de trabalho, sobre meu desempenho
78			Como parte do meu sistema de avaliação de desempenho, há uma discussão sobre qual treinamento preciso para melhorar meu desempenho
79			Não acredito que o sistema de avaliação de desempenho leve em consideração todas as minhas realizações profissionais
80			Meu gerente não conhece o suficiente sobre meu trabalho para me dar um resultado de avaliação de desempenho justo
81			O sistema que minha organização usa para alocar bônus/incentivos de avaliação de desempenho é justo
82			Como parte da minha última avaliação de desempenho, houve uma discussão sobre minha carreira e desenvolvimento pessoal
83			O processo de avaliação de desempenho da minha organização é uma avaliação justa do meu desempenho em relação a outros funcionários da minha organização
84			Sinto-me bastante satisfeito com as discussões sobre meu desempenho que tive com meu par
85			Estou satisfeito com o feedback que recebi do meu par
86			No geral, estou satisfeito com minhas interações com meu par
87			Sinto-me bastante satisfeito com as discussões sobre meu desempenho que tive com o diretor ou vice-diretor que me avaliou
88			Estou satisfeito com o feedback que recebi do diretor ou vice-diretor que me avaliou.
89			No geral, estou satisfeito com minhas interações com o diretor ou vice-diretor que me avaliou.
90			Em geral, sinto que a organização tem um excelente sistema de avaliação de desempenho
91			No geral, estou satisfeito com o sistema de avaliação de desempenho
92			A organização deve continuar a usar este sistema de avaliação de desempenho
93			Estou satisfeito com a forma como as avaliações de desempenho são tratadas
94			De modo geral, estou satisfeito com o processo de avaliação de desempenho
95	Utilidade	Quanto os colaboradores e os servidores percebem que o	Tenho uma ideia mais clara do que a organização espera por causa da minha avaliação
96			A avaliação me ajudou a aprender como posso melhorar meu trabalho

97		sistema de GD é proveitoso para eles e/ou para outras políticas, práticas e ações da organização.	Aprendi muito com a avaliação.
98			O processo de avaliação me ajudou a me desenvolver como profissional
99			A avaliação ajudou-me a aprender como posso fazer melhor o meu trabalho
100			Aprendi muito com a avaliação.
101			A avaliação me ajudou a entender meus erros
102			Tenho uma ideia mais clara do que o chefe espera de mim por causa da avaliação
103	Afetividade	Sentimentos experienciados pelos colaboradores e gestores em relação ao sistema de GD.	O processo de avaliação me causou muito estresse.
104			Eu estava tenso durante a revisão
105			A crítica me deixou com raiva
106			A entrevista foi perturbadora
107			Eu estava à vontade durante a maior parte da revisão

Instrumento B:

Dimensão Aprendizagem: Percepção dos colaboradores e gestores acerca de conhecimentos, habilidades e atitudes adquiridos a partir do sistema de GD.

Nº	Critério	Definição	Itens
1	Conhecimento	Informações declarativas, procedimentais ou tácitas adquiridas em decorrência da interação com o sistema de GD fornecendo clareza dos papéis desempenhados, das responsabilidades e dos resultados esperados.	Sei exatamente o que se espera de mim em relação à execução das tarefas
2			Tenho certeza de quantas responsabilidades eu tenho
3			Eu sei quais são minhas responsabilidades
4			Minhas tarefas são claras para mim
5			O sistema de GD esclareceu as expectativas em relação ao meu trabalho
6			O sistema de GD forneceu mais informações sobre minhas contribuições pessoais e valor agregado
7	Habilidade	Aplicação das informações adquiridas a partir do sistema de GD relacionados à habilidade de fornecer e receber feedback e de implementar adequadamente o sistema de GD.	Estou suficientemente treinado para executar minhas tarefas no que tange ao sistema de GD/AD
8			Não tenho as habilidades necessárias para executar tarefas no que tange ao sistema de GD/AD
9			Tenho o conhecimento necessário para executar minhas tarefas no que tange ao sistema de GD/AD
10			Sinto-me suficientemente competente para executar minhas tarefas no que tange ao sistema de GD/AD
11			Em todos os momentos, tenho a sensação de que dominei a execução das minhas tarefas no que tange ao sistema de GD/AD
12			Recebo feedback frequente e contínuo sobre como realizo meu trabalho
13			Recebo informações claras e diretas sobre meu desempenho no trabalho por meio de feedback fornecido continuamente
14			Raramente recebo feedback, exceto para sistemas formais de feedback, como avaliação de desempenho
15			No meu trabalho, sou continuamente informado sobre o que fiz bem ou o que poderia ter feito melhor
16			Sei pouco sobre o que meus colegas pensam sobre meu desempenho no trabalho
17			O sistema de AD/GD contribui para o meu desenvolvimento profissional
18	Atitude	Emoções e crenças demonstradas pelos colaboradores e gestores em relação ao sistema de GD.	Pretendo fazer mudanças em meu desempenho geral do trabalho com base em minha avaliação
19			Participo do sistema de GD/AD porque sou obrigado(a)
20			Tenho a intenção de melhorar meu desempenho baseado na avaliação que recebi
21			Eu acho que o sistema de GD/AD é interessante
22			Eu participo do sistema de GD/AD para o meu próprio bem

23			Eu participo do sistema de GD/AD porque eu deveria fazer isso
24			Pode haver boas razões para participar do sistema de GD/AD, mas pessoalmente não vejo nenhuma
25			Eu acho que o sistema de GD/AD é agradável
26			Eu acho que o sistema de GD/AD é bom para mim
27			Participar do sistema de GD/AD é algo que eu tenho que fazer
28			Eu participo do sistema de GD/AD, mas não tenho certeza se vale a pena
29			Participar do sistema de GD/AD é divertido
30			Participo do sistema de GD/AD por decisão pessoal
31			Participo do sistema de GD/AD porque eu não tenho escolha
32			Não sei, não vejo o que o sistema de GD/AD me traz
33			Eu me sinto bem ao participar do sistema de GD/AD
34			Eu acredito que o sistema de GD/AD é importante para mim
35			Eu sinto que eu tenho que fazer isso
36			Eu participo do sistema de GD/AD, mas não tenho certeza se é bom realizá-la

Instrumento C:

<u>Dimensão Transferência:</u> Comportamentos e resultados relacionados ao trabalho em si que são afetados por elementos do sistema de GD.			
Nº	Critério	Definição	Itens
1	Desempenho	Reconhecimento de melhoria nos comportamentos expressos no trabalho, voltados cada vez mais para o alcance de resultados organizacionais.	Quase sempre tenho um desempenho melhor do que o que pode ser caracterizado como desempenho aceitável
2			Frequentemente, tenho um desempenho melhor do que o esperado
3			Costumo fazer um esforço extra na realização do meu trabalho
4			Eu tento trabalhar o máximo possível
5			A qualidade do meu trabalho é excelente
6			Eu intencionalmente gasto muito esforço na realização do meu trabalho
7			Tenho desempenhado melhor minhas funções
8			Identifico efeitos claros no meu desempenho

9	Atitudes no trabalho	Sentimentos, crenças e percepções gerais do servidor e/ou supervisor sobre seu trabalho ou ambiente de trabalho relacionados à satisfação no trabalho, comprometimento organizacional e envolvimento com o trabalho.	Não me sinto “emocionalmente ligado” a esta organização
10			Não sinto um forte sentimento de pertença à minha organização
11			Esta organização tem um grande significado pessoal para mim
12			Eu realmente sinto como se os problemas desta organização fossem meus
13			Não me sinto ‘parte da família’ na minha organização
14			Eu gosto de discutir minha organização com pessoas de fora dela
15			Eu realmente sinto como se os problemas desta empresa fossem meus
16			No meu trabalho sinto-me cheio(a) de energia
17			Acho que o meu trabalho tem muito significado e utilidade
18			O tempo passa a voar quando estou a trabalhar
19			No meu trabalho sinto-me com força e energia
20			Estou entusiasmado(a) com o meu trabalho
21			Quando estou a trabalhar esqueço tudo o que se passa à minha roda
22			O meu trabalho inspira-me
23			Quando me levanto de manhã apetece-me ir trabalhar
24			Sinto-me feliz quando estou a trabalhar intensamente
25			Estou orgulhoso(a) do que faço neste trabalho
26			Estou imerso(a) no meu trabalho
27			Sou capaz de ficar a trabalhar por períodos muito longos
28			O meu trabalho é desafiante para mim
29			“Deixo-me ir” quando estou a trabalhar
30			Sou uma pessoa com muita resistência mental no meu trabalho
31			É difícil desligar-me do meu trabalho
32			No meu trabalho sou sempre perseverante (não desisto), mesmo quando as coisas não estão a correr bem
33			Eu geralmente sinto que meu trabalho vale a pena
34			Estou interessado no meu emprego atual
35			No geral, estou satisfeito com meu trabalho atual

36			Estou bastante satisfeito com o meu trabalho
37			Em geral, não gosto do meu trabalho.
38			Em suma, estou satisfeito com o meu trabalho.
39			Em geral, gosto de trabalhar aqui.
40	Motivação	Desejo dos colaboradores e dos gestores que compele a trabalhar de forma a alcançar os objetivos do trabalho.	As tarefas que realizo no trabalho são prazerosas
41			Meu trabalho é tão interessante que é uma motivação em si
42			As tarefas que realizo no trabalho representam, elas próprias, uma força motriz no meu trabalho
43			Meu trabalho é significativo
44			Eu me sinto sortudo por ser pago por um trabalho que eu gosto tanto
45			O trabalho é como um hobby para mim
46			Me motiva
47			Porque gosto muito deste trabalho
48			Porque eu me divirto fazendo meu trabalho
49			Pelos momentos de prazer que este trabalho me traz
50			Escolhi este trabalho porque me permite alcançar meus objetivos de vida
51			Porque este trabalho preenche meus planos de carreira
52			Porque este trabalho se encaixa nos meus valores pessoais
53			Porque tenho que ser o melhor no meu trabalho, tenho que ser um “vencedor”
54			Porque o meu trabalho é a minha vida e não quero falhar
55			Porque este trabalho me proporciona um certo padrão de vida
56			Porque minha reputação depende disso
57			Porque me permite ganhar muito dinheiro
58			Eu faço este trabalho pelo contracheque
59	Relação interpessoal	Interação social entre os indivíduos da organização é baseada em valores respeitosos	Recebo pedidos conflitantes de duas ou mais pessoas.
60			Não sei se meu trabalho será aceitável para meu chefe.
61			Meu gerente apoia meu crescimento e desenvolvimento profissional.
62			Meu gerente promove relacionamentos genuínos e de confiança na equipe.

63		e cordiais, fomentando relações de confiança.	Meu gerente me ajuda a obter os recursos de que preciso para fazer meu trabalho.
64	Bem-estar	Capacidade do trabalho em ser prazeroso e reduzir o sofrimento, promovendo a sensação de pertencimento e realização.	O trabalho me traz sensações positivas
65			O trabalho não me afeta negativamente na maior parte do tempo
66			O trabalho me auxilia a alcançar objetivos de vida

Informações sobre o sistema de GD/AD:

1. Qual sistema você acredita que é feito hoje na organização?

() Gestão de Desempenho

() Avaliação de desempenho

2. Os servidores são avaliados individualmente em qual(is) níveis:

() Autoavaliação

() Chefia imediata

() Integrantes da equipe de trabalho subordinada à chefia imediata

3. Há quanto tempo o sistema de GD/AD funciona na organização? _____

Informações sobre o sistema de GD/AD:

1. Em uma próxima etapa de pesquisa, vocês teriam os seguintes registro de dados individuais:

Produtividade individual: _____

Absenteísmo: _____

Rotatividade (interna): _____

Informações Sociodemográficas:

1. Qual sua idade? _____

2. Sexo:

() Masculino

() Feminino

() Prefiro não identificar

3. Por favor, indique sua escolaridade completa:

() Ensino Fundamental

() Ensino Médio

() Ensino Superior

() Pós-graduação/MBA/Especialização

() Mestre

() Doutor

Apêndice C – Instrumento 2 – Validação de juízes

Instrumento de Reação

ORIENTAÇÕES PARA VALIDAÇÃO DO INSTRUMENTO - REAÇÃO

Prezado(a) avaliador(a),

Este documento compreende o processo de construção e validação do instrumento para avaliação da efetividade do sistema de gestão de desempenho* - como sendo um sistema estruturante composto por práticas, políticas e procedimentos responsável por, a partir da avaliação do desempenho, promover o desempenho individual de forma consistente com os objetivos estratégicos, visando o alcance de desempenho organizacional superior - no setor público.

* *Gestão de Desempenho (GD) e a Avaliação de Desempenho (AD) devem ser tratados de forma intercambiável na análise do instrumento.*

Esta escala refere-se à medida **Reação** (aba na planilha), a qual possui **58 itens**, que tem como objetivo mensurar uma das dimensões da efetividade dos sistemas de GD no setor público. Os itens foram gerados a partir de traduções livres de instrumentos validados no exterior, tanto quanto de entrevistas realizadas com servidores públicos de órgãos variados.

Neste sentido, foi elaborada a seguinte definição:

Definição de Reação : Impressões, julgamentos, avaliações e opiniões em geral dos colaboradores e gestores sobre o sistema de GD após a participação ou contato com ações ou práticas relacionadas ao sistema de GD.

A **Reação** em relação aos sistemas de GD/AD é expressa por **4 fatores: Afetividade, Justiça, Precisão e Utilidade**. Estes quatro fatores são determinados da seguinte forma**:

i. Afetividade: Sentimentos experienciados pelos servidores em relação ao sistema de GD.

ii. Justiça: Quanto os servidores percebem que o sistema de GD são consistentes com as normas, garantindo equidade e igualdade; que o processo adere a critérios de consistência, ausência de vieses, representação, precisão e ética e permite influenciar ou participar da tomada de decisão; a relação entre o avaliado e o avaliador é respeitosa e sensível, com clareza nos processos de tomada de decisão.

iii. Precisão: Quanto os servidores percebem que o sistema de GD é preciso e confiável em relação ao planejamento do desempenho, alinhando metas individuais e organizacionais; ao feedback e aconselhamento contínuos; à mensuração do desempenho atual e o esperado; e aos resultados (gratificações, benefícios, capacitação etc.).

iv. Utilidade: Quanto os servidores percebem que o sistema de GD é proveitoso para eles e/ou para outras políticas, práticas e ações da organização.

**** As definições de cada um dos fatores também estão contidas nas abas dos seus respectivos instrumentos.**

Sua tarefa consiste em analisar dois aspectos de cada item:

a) Adequação do item ao fator: verificação de a qual faceta pertence cada item, tendo como base a definição apresentada. Para realizar esta análise de conteúdo, assinala com um "X" ao fator que você considera estar representado pela afirmação do item analisado. Se considerar que o conteúdo do item não apresenta relação com as facetas apresentadas, deixe-o em branco (não marque nada).

b) Clareza: avaliação do quanto esses itens são compreensíveis (diretos, claros e objetivos). Para esta avaliação, assinala com um "X" se o item em análise é compreensível ou incompreensível. No caso de ser incompreensível, você poderá sugerir alguma adequação, caso queira. Sugestões são bem-vindas e devem ser feitas no próprio instrumento de avaliação.

Com este instrumento você também está recebendo um termo de consentimento livre e esclarecido por e-mail. Pedimos a gentileza de que devolva junto com a validação. Pedimos, ainda, sua compreensão em não divulgar os itens deste instrumento, por se tratar de um teste que terá caráter sigiloso.

Muito obrigada pela contribuição.

Bom trabalho.

Definição de Reação: Impressões, julgamentos, avaliações e opiniões em geral dos colaboradores e gestores sobre o sistema de GD após a participação ou contato com ações ou práticas relacionadas ao sistema de GD.

i. Afetividade: Sentimentos experienciados pelos servidores em relação ao sistema de GD.

ii. Justiça: Quanto os servidores percebem que o sistema de GD são consistentes com as normas, garantindo equidade e igualdade; que o processo adere a critérios de consistência, ausência de vieses, representação, precisão e ética e permite influenciar ou participar da tomada de decisão; a relação entre o avaliado e o avaliador é respeitosa e sensível, com clareza nos processos de tomada de decisão.

iii. Precisão: Quanto os servidores percebem que o sistema de GD é preciso e confiável em relação ao planejamento do desempenho, alinhando metas individuais e organizacionais; ao feedback e aconselhamento contínuos; à mensuração do desempenho atual e o esperado; e aos resultados (gratificações, benefícios, capacitação etc.).

iv. Utilidade: Quanto os servidores percebem que o sistema de GD é proveitoso para eles e/ou para outras políticas, práticas e ações da organização

Observação: Itens com * são reversos.

Pedimos que pensem no item em resposta a seguinte frase: "A partir da minha interação com o sistema de GD,..."

nt	Itens	Análise:	Análise de Conteúdo (Adequação do item ao fator)				Análise Semântica (Clareza)						
		Orientação:	Marque um X no fator a que se refere o item. Assinale apenas 1 (um) fator				Atribuir nota entre 1 a 5 em relação à compreensão do item. (Caso marque 1 ou 2, reescreva o item indicando as melhorias)						
		Fator:	Afetividade	Justiça	Precisão	Utilidade	Incompreensível	1	2	3	4	5	Compreensível
A partir da minha interação com o sistema de Gestão do Desempenho...													
1	Sinto que a avaliação que recebo do meu desempenho é consistente com o que eu mereço												
2	Acho a oportunidade ter meu desempenho avaliado												
3	Acredito que a avaliação de desempenho me ajuda a aprender como posso fazer melhor o meu trabalho												
4	Acredito que a avaliação de desempenho me auxilia a emergir chances de crescimento na organização												
5	Acredito que a avaliação de desempenho que recebo corresponde a minha autoavaliação												
6	Acredito que a avaliação de desempenho que recebo considera objetivamente o meu desempenho atual em relação ao esperado												
7	Acredito que a avaliação de desempenho realizada representa uma aferição precisa do meu desempenho												
8	Acredito que a avaliação de desempenho respeita padrões éticos e morais												
9	Acredito que a avaliação do meu desempenho me ajuda a entender meus erros												
10	Acredito que a avaliação que recebo do meu desempenho é justa												
11	Acredito que a avaliação que recebo do meu desempenho fornece uma avaliação adequada do trabalho que realizo												
12	Acredito que a avaliação que recebo do meu desempenho me fornece uma boa avaliação do esforço que coloco no meu trabalho												
13	Acredito que a minha chefia imediata avalia meu desempenho de forma educada												
14	Acredito que as minhas metas ou objetivos são precisamente avaliados como parte do processo de avaliação do desempenho												
15	Acredito que eu aprendo muito com a avaliação de desempenho que recebo												
16	Acredito que minha chefia imediata avalia meu desempenho de forma justa												
17	Acredito que recebo feedback construtivo sobre o meu desempenho no trabalho por parte do(s) meu(s) avaliador(es)												
18	Acredito que recebo feedback oportuno sobre o meu desempenho no trabalho por parte da minha chefia imediata												
19	Acredito que as minhas recompensas estão relacionadas à avaliação do desempenho que recebi												
20	Compreendo as razões da avaliação de desempenho que recebo												
21	Compreendo o que devo fazer para receber uma avaliação de desempenho melhor												
22	Compreendo que o feedback contínuo me fornece uma avaliação precisa de como estou me saindo em relação ao desempenho esperado												
23	Concentro melhor meus esforços a partir da identificação das metas relevantes para o atingimento dos objetivos da organização												
24	Conflito que a minha chefia imediata comunica claramente o desempenho esperado												
25	Conheço as motivações utilizadas pela minha chefia imediata para avaliar meu desempenho												
26	Considero importante minha avaliação de desempenho ser atrelada a remuneração que recebo												
27	Considero que a avaliação de desempenho que recebo é livre de vieses												
28	Considero que a avaliação de desempenho que recebo é precisa em relação ao meu desempenho												
29	Considero que a avaliação do meu desempenho reflete o meu verdadeiro desempenho												
30	Considero que a avaliação que recebo do meu desempenho afere o que tenho contribuído para a organização												
31	Considero que a minha chefia imediata avalia meu desempenho com dignidade												
32	Considero que a minha chefia imediata avalia o desempenho dos funcionários de forma justa												
33	Considero que a minha chefia imediata se abstém de observações ou comentários impróprios ao avaliar meu desempenho												
34	Considero que a periodicidade com que a avaliação de desempenho é aplicada é apropriada para capturar o meu desempenho de forma												
35	Considero que o feedback anual que recebo é uma representação precisa dos feedbacks contínuos recebidos ao longo daquele ciclo de												
36	Considero que o processo de avaliação do desempenho me ajuda a me desenvolver como profissional												
37	Consigo visualizar como a minha avaliação de desempenho se compara a das outras pessoas em minha organização que realizam atividades												
38	Estou de acordo com a avaliação final que recebo do meu desempenho												
39	Fico à vontade ao ter meu desempenho avaliado												
40	Fico feliz em receber o resultado da minha avaliação de desempenho												
41	Fico tenso(a) ao ter meu desempenho avaliado*												
42	Identifico que são utilizadas medidas objetivas para avaliar meu desempenho												
43	Minha chefia imediata e eu atualizamos minhas metas conforme os objetivos da organização mudam												
44	Participo do estabelecimento dos indicadores que são utilizados para avaliar meu desempenho no trabalho												
45	Percebo que a avaliação de desempenho me fez refletir sobre o meu desempenho												
46	Percebo que as recompensas que recebo são baseadas no meu desempenho												
47	Posso recorrer da avaliação de desempenho que recebo caso eu não concorde												
48	Recebo orientação ou assistência para me ajudar a melhorar meu desempenho por parte da minha chefia imediata												
49	Sinto que a avaliação de desempenho me causa estresse*												
50	Sinto que a avaliação de desempenho na organização é utilizada para além de objetivos administrativos e profoma												
51	Sinto que a avaliação que recebo é justificada dado o meu desempenho												
52	Sinto que durante o período de avaliação, minhas possibilidades de melhoria são claramente apontadas para mim												
53	Sinto que exerce influência no processo de avaliação de desempenho da minha organização												
54	Sinto riva ao ler meu desempenho avaliado*												
55	Sinto-me confortável em expressar minhas opiniões e sentimentos sobre o processo de avaliação de desempenho da minha organização												
56	Sinto-me informado pela minha chefia imediata sobre o quão bem estou desempenhando minhas atividades												
57	Tenho oportunidades suficientes (como atividades ou projetos desafiadores) para obter uma avaliação de desempenho melhor												
58	Tenho receio sobre o que a avaliação de desempenho pode indicar*												

Instrumento de Aprendizagem

ORIENTAÇÕES PARA VALIDAÇÃO DO INSTRUMENTO - APRENDIZAGEM

Prezado(a) avaliador(a),

Este documento compreende o processo de construção e validação do instrumento para avaliação da efetividade do sistema de gestão de desempenho* - como sendo um sistema estruturante composto por práticas, políticas e procedimentos responsável por, a partir da avaliação do desempenho, promover o desempenho individual de forma consistente com os objetivos estratégicos, visando o alcance de desempenho organizacional superior - no setor público.

** Gestão de Desempenho (GD) e a Avaliação de Desempenho (AD) devem ser tratadas de forma intercambiável na análise do instrumento.*

Esta escala refere-se à medida **Aprendizagem**, a qual possui **34** itens, que tem como objetivo mensurar uma das dimensões da efetividade dos sistemas de GD no setor público. Os itens foram gerados a partir de traduções livres de instrumentos validados no exterior, tanto quanto de entrevistas realizadas com servidores públicos de órgãos variados.

Neste sentido, foi elaborada a seguinte definição:

Definição de Aprendizagem: **Percepção dos colaboradores e gestores acerca de conhecimentos, habilidades e atitudes adquiridos a partir do sistema de GD.**

A **Aprendizagem** em relação aos sistemas de GD é expressa por 3 fatores: **Atitude, Conhecimento e Habilidade**. Estes três fatores são determinados da seguinte forma**:

i. **Atitude:** Emoções e crenças demonstradas pelos colaboradores e gestores em relação ao sistema de GD.

ii. **Conhecimento:** Informações declarativas, procedimentais ou tácitas adquiridas em decorrência da interação com o sistema de GD fornecendo clareza dos papéis desempenhados, das responsabilidades e dos resultados esperados.

iii. **Habilidade:** Aplicação das informações adquiridas a partir do sistema de GD relacionados à habilidade de fornecer e receber feedback e de implementar adequadamente o sistema de GD.

**** As definições de cada um dos fatores também estão contidas nas abas dos seus respectivos instrumentos.**

Sua tarefa consiste em analisar dois aspectos de cada item:

a) Adequação do item ao fator: verificação de a qual faceta pertence cada item, tendo como base a definição apresentada. Para realizar esta análise de conteúdo, assinale com um "X" ao fator que você considera estar representado pela afirmação do item analisado. Se considerar que o conteúdo do item não apresenta relação com as facetas apresentadas, deixe-o em branco (não marque nada).

b) Clareza: avaliação do quanto esses itens são compreensíveis (diretos, claros e objetivos). Para esta avaliação, assinale com um "X" se o item em análise é compreensível ou incompreensível. No caso de ser incompreensível, você poderá sugerir alguma adequação, caso queira. Sugestões são bem-vindas e devem ser feitas no próprio instrumento de avaliação.

Com este instrumento você também está recebendo um termo de consentimento livre e esclarecido por e-mail. Pedimos a gentileza de que devolva junto com a validação. Pedimos, ainda, sua compreensão em não divulgar os itens deste instrumento, por se tratar de um teste que terá caráter sigiloso.

Muito obrigada pela contribuição.

Bom trabalho.

Definição de **Aprendizagem**: Percepção dos colaboradores e gestores acerca de conhecimentos, habilidades e atitudes adquiridos a partir do sistema de GD.

i. **Atitude:** Emoções e crenças demonstradas pelos colaboradores e gestores em relação ao sistema de GD.

ii. **Conhecimento:** Informações declarativas, procedimentais ou tácitas adquiridas em decorrência da interação com o sistema de GD fornecendo clareza dos papéis desempenhados, das responsabilidades e dos resultados esperados.

iii. **Habilidade:** Aplicação das informações adquiridas a partir do sistema de GD relacionados à habilidade de fornecer e receber feedback e de implementar adequadamente o sistema de GD.

Observação: Itens com * são reversos.

Pedimos que pensem o item em resposta a seguinte frase: **"A partir da minha interação com o sistema de GD,..."**

nº	Itens	Análise:	Análise de Conteúdo (Adequação do item ao fator)			Análise Semântica
		Orientação:	Marque um X no fator a que se refere o item. (Assinale apenas 1 (um) fator)			Caso marque 1 ou 2, reescreva o item indicando as melhorias
		Fator:	Atitude	Conhecimento	Habilidade	Incompreensível 1 2 3 4 5 Compreensível
A partir da minha interação com o sistema de Gestão do Desempenho...						
1	Considero que a avaliação de desempenho contribui para o meu desenvolvimento profissional					
2	Acho que a avaliação de desempenho é interessante					
3	Acho que o sistema de GD/AD é bom para mim					
4	Acredito que a avaliação de desempenho esclarece as expectativas em relação ao meu trabalho					
5	Acredito que a avaliação de desempenho fornece informações sobre minhas contribuições pessoais e meu valor na organização					
6	Adquiro o conhecimento necessário para executar minhas tarefas relacionadas ao meu papel na avaliação de desempenho					
7	Considero que a avaliação de desempenho é importante para mim					
8	Considero que o processo de avaliação de desempenho é agradável					
9	Considero que participar da avaliação de desempenho é algo que eu tenho que fazer					
10	Considero que participar da avaliação de desempenho é divertido					
11	Considero-me competente para executar minhas tarefas relacionadas ao meu papel na avaliação de desempenho					
12	Consigo receber ou fornecer informações claras e diretas sobre o desempenho esperado					
13	Domino melhor as tarefas relacionadas ao meu papel na avaliação de desempenho					
14	Entendo as razões pelas quais eu devo participar da avaliação de desempenho					
15	Entendo que pode haver boas razões para contribuir com a avaliação de desempenho, mas pessoalmente não vejo nenhuma*					
16	Minhas tarefas ficam claras para mim					
17	Obtenho as habilidades necessárias para executar as tarefas relacionadas ao meu papel na avaliação de desempenho					
18	Participo da avaliação de desempenho para o meu próprio bem					
19	Sei como receber ou fornecer feedback mesmo se não forem momentos formais de feedback ou de avaliação do desempenho					
20	Sei o que se espera de mim em relação à execução das tarefas					
21	Sei quais são minhas responsabilidades					
22	Sinto que tenho que participar da avaliação de desempenho					
23	Sinto-me bem ao participar da avaliação de desempenho					
24	Sinto-me capaz de fazer mudanças em meu desempenho geral do trabalho baseadas na minha avaliação					
25	Sinto-me capaz de receber ou fornecer feedback frequente e contínuo sobre o trabalho realizado					
26	Sinto-me capaz de reconhecer o que fiz bem e o que poderia ter feito melhor no trabalho					
27	Sinto-me disposto(a) a participar da avaliação de desempenho por decisão pessoal					
28	Sinto-me disposto(a) a participar da avaliação de desempenho porque eu não tenho escolha					
29	Sinto-me disposto(a) a participar da avaliação de desempenho porque sou obrigado(a)					
30	Sinto-me disposto(a) a participar da avaliação de desempenho, mas não tenho certeza se é bom realizá-la*					
31	Sinto-me disposto(a) a participar da avaliação de desempenho, mas não tenho certeza se vale a pena*					
32	Sinto-me treinado para executar minhas atribuições relacionadas ao meu papel na avaliação de desempenho					
33	Tenho a intenção de melhorar meu desempenho baseado na avaliação que recebi					
34	Vejo os benefícios que a avaliação de desempenho me traz					

Instrumento de Transferência

ORIENTAÇÕES PARA VALIDAÇÃO DO INSTRUMENTO - TRANSFERÊNCIA

Prezado(a) avaliador(a),

Este documento compreende o processo de construção e validação do instrumento para avaliação da efetividade do sistema de gestão de desempenho* - como sendo um sistema estruturante composto por práticas, políticas e procedimentos responsável por, a partir da avaliação do desempenho, promover o desempenho individual de forma consistente com os objetivos estratégicos, visando o alcance de desempenho organizacional superior - no setor público.

** Gestão de Desempenho (GD) e a Avaliação de Desempenho (AD) devem ser tratadas de forma intercambiável na análise do instrumento.*

Esta escala refere-se à medida **Transferência**, a qual possui **56** itens, que tem como objetivo mensurar uma das dimensões da efetividade dos sistemas de GD no setor público. Os itens foram gerados a partir de traduções livres de instrumentos validados no exterior, tanto quanto de entrevistas realizadas com servidores públicos de órgãos variados.

Neste sentido, foi elaborada a seguinte definição:

Definição de Transferência: **Comportamentos e resultados relacionados ao trabalho em si que são afetados por elementos do sistema de GD.**

A **Transferência** em relação aos sistemas de GD/AD é expressa por **5** fatores: **Atitudes no trabalho, Bem-estar, Desempenho, Motivação e Relação interpessoal**. Estes quatro fatores são determinados da seguinte forma**:

i. **Atitudes no trabalho:** Sentimentos, crenças e percepções gerais do servidor e/ou supervisor sobre seu trabalho ou ambiente de trabalho relacionados à satisfação no trabalho, comprometimento organizacional e envolvimento com o trabalho.

ii. **Bem-estar:** Capacidade do trabalho em ser prazeroso e reduzir o sofrimento, promovendo a sensação de pertencimento e realização.

iii. **Desempenho:** Reconhecimento de melhoria nos comportamentos expressos no trabalho, voltados cada vez mais para o alcance de resultados organizacionais.

iv. **Motivação:** Desejo dos colaboradores e dos gestores que compele a trabalhar de forma a alcançar os objetivos do trabalho.

v. **Relação interpessoal:** Interação social entre os indivíduos da organização é baseada em valores respeitosos e cordiais, fomentando relações de confiança.

Sua tarefa consiste em analisar dois aspectos de cada item:

a) Adequação do item ao fator: verificação de a qual faceta pertence cada item, tendo como base a definição apresentada. Para realizar esta análise de conteúdo, assinale com um "X" ao fator que você considera estar representado pela afirmação do item analisado. Se considerar que o conteúdo do item não apresenta relação com as facetas apresentadas, deixe-o em branco (não marque nada).

b) Clareza: avaliação do quanto esses itens são compreensíveis (diretos, claros e objetivos). Para esta avaliação, assinale com um "X" se o item em análise é compreensível ou incompreensível. No caso de ser incompreensível, você poderá sugerir alguma adequação, caso queira. Sugestões são bem-vindas e devem ser feitas no próprio instrumento de avaliação.

Com este instrumento você também está recebendo um termo de consentimento livre e esclarecido por e-mail. Pedimos a gentileza de que devolva junto com a validação. Pedimos, ainda, sua compreensão em não divulgar os itens deste instrumento, por se tratar de um teste que terá caráter sigiloso.

Muito obrigada pela contribuição.

Bom trabalho.

Dimensão Transferência

Comportamentos e resultados relacionados ao trabalho em si que são afetados por elementos do sistema de GD.

Atitudes no trabalho

Sentimentos, crenças e percepções gerais do servidor e/ou supervisor sobre seu trabalho ou ambiente de trabalho relacionados à satisfação no trabalho, comprometimento organizacional e envolvimento com o trabalho.

Bem-estar

Capacidade do trabalho em ser prazeroso e reduzir o sofrimento, promovendo a sensação de pertencimento e realização.

Desempenho

Reconhecimento de melhoria nos comportamentos expressos no trabalho, voltados cada vez mais para o alcance de resultados organizacionais.

Motivação

Desejo dos colaboradores e dos gestores que compete a trabalhar de forma a alcançar os objetivos do trabalho.

Relação interpessoal

Interação social entre os indivíduos da organização é baseada em valores respeitosos e cordais, fomentando relações de confiança.

Podemos que pensem no item em resposta a seguinte frase: "A partir da minha interação com o sistema de GD..."

nº	Item	Análise:	Análise de Conteúdo					Análise Semântica
		Orientação:	(Adequação do item ao fator)					
			(Marque um X no fator a que se refere o item. (Assinale apenas 1 (um) fator))					
		Fator:	Atitudes no trabalho	Bem-estar	Desempenho	Motivação	Relação Interpessoal	Caso marque 1 ou 2, reescreva o item indicando as melhorias
A partir da minha interação com o sistema de Gestão do Desempenho...								
1	Acredito que o tempo via quando estou trabalhando							
2	Considero a qualidade do meu trabalho excelente							
3	Considero meu trabalho uma motivação em si							
4	Considero que a minha equipe é feita por relacionamentos genuínos							
5	Considero que as tarefas que realizo no trabalho representam, elas próprias, uma força motriz no meu trabalho							
6	Considero que as tarefas que realizo no trabalho são prazerosas							
7	Considero que este trabalho preenche meus planos de carreira							
8	Considero que este trabalho se encaixa nos meus valores pessoais							
9	Considero que o meu trabalho é inspirador							
10	Considero que o meu trabalho é útil							
11	Considero que o meu trabalho me motiva							
12	Considero que o meu trabalho me permite ganhar muito dinheiro							
13	Considero que o trabalho é como um hobby para mim							
14	Considero que os problemas desta organização são meus							
15	Desempenho melhor minhas funções							
16	Divirto-me fazendo meu trabalho							
17	Empenho-me na realização do meu trabalho							
18	Esqueço tudo o que se passa a minha volta quando estou trabalhando							
19	Estou bastante satisfeito(a) com o meu trabalho							
20	Estou entusiasmado(a) com o meu trabalho							
21	Estou imerso(a) no meu trabalho							
22	Estou interessado(a) no meu emprego atual							
23	Estou orgulhoso(a) do que faço neste trabalho							
24	Faço este trabalho pelo contracheque							
25	Faço um esforço extra na realização do meu trabalho							
26	Gosto de conversar sobre a minha organização com pessoas de fora dela							
27	Gosto de trabalhar aqui							
28	Gosto muito do meu trabalho							
29	Identifico efeitos claros no meu desempenho							
30	Posso receber elogios e pedidos mais claros mesmo que sejam de duas ou mais pessoas							
31	Sinto dificuldade em me desligar do trabalho							
32	Sinto que "deixo-me ir" quando estou trabalhando							
33	Sinto que a minha reputação depende do meu trabalho							
34	Sinto que a organização representa um grande significado pessoal para mim							
35	Sinto que este trabalho me traz momentos de prazer							
36	Sinto que meu crescimento e desenvolvimento profissional são apoiados							
37	Sinto que meu trabalho vale a pena							
38	Sinto que o meu trabalho é a minha vida							
39	Sinto que o meu trabalho é desafiante para mim							
40	Sinto que o meu trabalho me proporciona um certo padrão de vida							
41	Sinto que tenho que "dar" o melhor no meu trabalho							
42	Sinto que trabalhar é instigante							
43	Sinto sensações positivas a partir do meu trabalho							
44	Sinto um forte sentimento de pertencimento à minha organização							
45	Sinto-me "parte da família" na minha organização							
46	Sinto-me "emocionalmente ligado(a)" a esta organização							
47	Sinto-me capaz de trabalhar por períodos muito longos							
48	Sinto-me cheio(a) de energia quando estou trabalhando							
49	Sinto-me feliz quando trabalho intensamente							
50	Sinto-me sortudo(a) por ser pago por um trabalho que eu gosto tanto							
51	Sou auxiliado(a) em relação a obter os recursos de que preciso para fazer meu trabalho							
52	Sou persistente no trabalho, mesmo quando as coisas não estão a correr bem							
53	Tenho clareza sobre o quão aceitável meu trabalho é para o meu chefe							
54	Tenho um desempenho melhor do que o esperado							
55	Tenho um desempenho melhor do que o que pode ser caracterizado como desempenho aceitável							
56	Tenho trabalhar o máximo possível							

Apêndice D – Questionário AFE

Pesquisa sobre Efetividade da Gestão de Desempenho no Setor Público

Prezado(a) Participante,

Você está sendo convidado(a) a participar de um estudo de pesquisa acadêmica sobre a efetividade da Gestão de Desempenho (GD) ou Avaliação de Desempenho (AD) no setor público conduzido pela pesquisadora Mariana Lopes de Araújo, doutoranda do Programa de Pós-Graduação em Administração da Universidade de Brasília (UnB), orientada pelo Professor Doutor Pedro Menezes.

Antes de prosseguir pedimos que leia atentamente as informações a seguir e que compreenda completamente o que a participação nesta pesquisa implica. Sua participação é voluntária e você tem o direito de se recusar a participar ou desistir do estudo a qualquer momento, sem sofrer qualquer consequência.

Esta pesquisa tem como objetivo **testar empiricamente um modelo de avaliação da efetividade do sistema de Gestão de Desempenho adaptado ao contexto da Administração Pública brasileira.**

Se você concordar em participar deste estudo, será solicitado(a) que você responda ao questionário a seguir sobre sua experiência com o processo de GD/AD em organizações públicas em relação a uma das três dimensões: Reação, Aprendizagem ou Transferência. Lembre-se de que não há respostas certas ou erradas. Todas as respostas são aceitas desde que representem sua opinião. **O tempo estimado para conclusão deste questionário é de aproximadamente 10 minutos.**

As suas respostas e informações pessoais coletadas neste questionário serão tratadas de forma confidencial. Todos os dados serão tratados de maneira agregada e anônima, sem a identificação individual dos participantes. Os resultados serão usados apenas para fins de pesquisa e serão apresentados de forma geral, sem mencionar nomes ou qualquer informação pessoal.

Não há riscos significativos associados à participação neste estudo. Ao responder você contribuirá para o desenvolvimento de uma medida de efetividade do sistema de Gestão de Desempenho. Lembre-se de que não há respostas certas ou erradas.

Se você tiver alguma dúvida ou preocupação em relação a esta pesquisa, entre em contato com Mariana Lopes pelo e-mail 200047167@aluno.unb.br.

Agradecemos sinceramente sua participação nesta pesquisa.
Atenciosamente,

Msc. Mariana Lopes de Araújo

Doutoranda em Administração da Universidade de Brasília (UnB).

* Obrigatória

1. Se você concorda em participar deste estudo, clique em "Concordo" abaixo para indicar seu consentimento. Caso contrário, clique em "Não concordo" e sua participação não será registrada. *

☐ Concordo

☐ Não concordo

2. **Você já participou ou participa, seja como avaliado(a) ou avaliador(a), de um processo formal de Avaliação de Desempenho (AD)/Gestão de Desempenho (GD) em algum órgão público? ***

☐ Sim

☐ Não

3. **Escolha o período do ano em que você faz aniversário: ***

Esta informação é apenas para triagem dos questionários e não tem relação com o tema da pesquisa.

☐ 01 de janeiro a 30 de abril

☐ 01 de maio a 31 de agosto

☐ 01 de setembro a 31 de dezembro

REAÇÃO

Esta seção do questionário visa avaliar:

- **Reação:** As impressões, julgamentos, avaliações e opiniões em geral dos servidores sobre a GD/AD após a participação ou contato com ações ou práticas relacionadas à GD/AD.

Pedimos que avalie cada uma das afirmativas a seguir de acordo com o quanto você concorda ou discorda de cada uma delas.

Recomendamos que você leia cada pergunta atentamente e responda:

A partir da minha interação com a Gestão do Desempenho (GD)/Avaliação de Desempenho (AD), percebo que...

4. Fico feliz em receber o resultado da minha avaliação de desempenho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

5. Fico tenso(a) ao ter meu desempenho avaliado *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

6. A avaliação de desempenho me causa estresse *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

7. Sinto raiva ao ter meu desempenho avaliado *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

8. A avaliação de desempenho que recebo respeita os padrões éticos e morais *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

9. A avaliação que recebo do meu desempenho é equânime *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

10. A avaliação que recebo do meu desempenho me fornece uma avaliação adequada do trabalho que eu realizo *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

11. A avaliação que recebo do meu desempenho me fornece uma boa avaliação do esforço que coloco no meu trabalho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

12. Minha chefia imediata avalia o desempenho dos funcionários de forma justa *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

13. A minha participação é considerada no estabelecimento dos indicadores que são utilizados para avaliar meu desempenho no trabalho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

14. As recompensas que recebo são baseadas no meu desempenho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

15. Obtenho informações sobre o quão bem estou desempenhando minhas atividades sob o ponto de vista da minha chefia *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

16. A avaliação de desempenho que recebo é precisa em relação ao meu desempenho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

17. A avaliação do meu desempenho reflete o meu verdadeiro desempenho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

18. Minha chefia imediata e eu atualizamos minhas metas conforme os objetivos da organização mudam *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

19. Me ajuda a enxergar chances de crescimento na organização *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

20. Aprendo muito com a avaliação de desempenho que recebo *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

21. O processo de avaliação do desempenho me ajuda a me desenvolver como profissional *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

22. Me faz refletir sobre o meu desempenho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

APRENDIZAGEM

Este questionário visa avaliar:

- **Aprendizagem:** Conhecimentos, habilidades e atitudes que os servidores percebem ter adquirido a partir da GD/AD.

Pedimos que avalie cada uma das afirmativas a seguir de acordo com o quanto você concorda ou discorda de cada uma delas.

Recomendamos que você leia cada pergunta atentamente e responda:

A partir da minha interação com a Gestão do Desempenho (GD)/Avaliação de Desempenho (AD), percebo que...

23. O processo de GD/AD é agradável *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

24. Me traz benefícios *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

25. Participar da GD/AD é divertido *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

26. Participo da GD/AD para o meu próprio bem *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

27. Devo continuar participando da GD/AD *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

28. Me sinto bem ao participar da GD/AD *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

29. Estou disposto(a) a participar da GD/AD por decisão pessoal *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

30. Participar da GD/AD é algo que eu tenho que fazer *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

31. Estou disposto(a) a participar da GD/AD porque eu não tenho escolha *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

32. Estou disposto(a) a participar da GD/AD porque sou obrigado(a) *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

33. Estou disposto(a) a participar da GD/AD, mas não tenho certeza se é bom realizá-la *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

34. Estou disposto(a) a participar da GD/AD, mas não tenho certeza se vale a pena* *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

35. As expectativas em relação ao meu trabalho são esclarecidas *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

36. A GD/AD fornece informações sobre minhas contribuições pessoais e meu valor na organização *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

37. Minhas tarefas ficam claras para mim *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

38. Obtenho as habilidades necessárias para executar as tarefas relacionadas ao meu papel na GD/AD *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

39. Conheço melhor o que se espera de mim em relação à execução das tarefas que devo realizar *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

40. Posso executar minhas tarefas relacionadas ao meu papel na GD/AD *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

41. Posso receber ou fornecer informações claras e diretas sobre o meu desempenho esperado *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo totalmente

42. Sei como receber ou fornecer feedback mesmo se não forem momentos formais de devolutiva ou avaliativos *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo totalmente

43. Sou capaz de fazer mudanças em meu desempenho a partir da avaliação que recebo *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo totalmente

44. Posso tranquilamente receber ou fornecer feedback sobre o trabalho realizado *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo totalmente

45. Me sinto suficientemente capaz de desempenhar meu papel como avaliado(a) ou avaliador(a) na GD/AD *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo totalmente

TRANSFERÊNCIA

Esta seção do questionário visa avaliar

- **Transferência: Comportamentos e resultados relacionados ao trabalho em si que são afetados pela GD/AD.**

Pedimos que avalie cada uma das afirmativas a seguir de acordo com o quanto você concorda ou discorda de cada uma delas.

Recomendamos que você leia cada pergunta atentamente e responda:

A partir da minha interação com a Gestão do Desempenho (GD)/Avaliação de Desempenho (AD), percebo que...

46. O tempo voa quando estou trabalhando *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

47. Meu trabalho é útil *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

48. Estou satisfeito(a) com o meu trabalho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

49. Estou imerso(a) no meu trabalho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

50. Estou interessado(a) no meu emprego atual *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

51. Estou orgulhoso(a) do que faço neste trabalho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

52. Gosto de conversar sobre a minha organização com pessoas de fora dela *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

53. Gosto de trabalhar na minha organização *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

54. Gosto muito do meu trabalho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

55. "Me deixo ir" quando estou trabalhando *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

56. A organização tem um grande significado pessoal para mim *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

57. Meu trabalho vale a pena *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

58. Meu trabalho gera sensações positivas *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

59. A qualidade do meu trabalho é excelente *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

60. Meu desempenho está melhor do que o esperado *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

61. Meu desempenho está melhor do que o que é considerado um
desempenho aceitável *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

62. Meu trabalho é uma motivação por si só *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

63. Este trabalho atende aos meus planos de carreira *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

64. Meu trabalho me motiva *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

65. Meus colegas e chefia me auxiliam no que preciso para fazer meu trabalho *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

66. Obtenho clareza sobre o quão aceitável meu trabalho é para a minha chefia *

1	2	3	4	5
---	---	---	---	---

Discordo totalmente

Concordo
totalmente

Dados para caracterização da amostra

Assinale a opção correspondente.

67. Você é vinculado(a) a qual esfera governamental? *

- ☐ Federal
- ☐ Estadual
- ☐ Municipal
- ☐ Distrital

68. Você é vinculado(a) de qual dos poderes? *

- ☐ Executivo
- ☐ Legislativo
- ☐ Judiciário
- ☐ Órgão Independente (Ex. Tribunal de Contas)

69. Gênero *

- ☐ Feminino
- ☐ Masculino
- ☐ Outro
- ☐ Prefiro não responder

70. Idade *

- ☐ 18 a 24 anos
- ☐ 25 a 34 anos
- ☐ 35 a 44 anos
- ☐ 45 a 55 anos
- ☐ 55 a 64 anos
- ☐ 65 anos ou mais
- ☐ Prefiro não responder

71. Escolaridade (maior nível) *

- ☐ Ensino Fundamental
- ☐ Ensino Médio
- ☐ Ensino Superior
- ☐ Especialização
- ☐ Mestrado
- ☐ Doutorado

72. Faixa salarial *

- ☐ Até um salário mínimo
- ☐ De 2 a 4 salários mínimos
- ☐ De 5 a 7 salários mínimos
- ☐ De 8 a 10 salários mínimos
- ☐ Acima de 10 salários mínimos
- ☐ Prefiro não responder

73. Digite seu e-mail abaixo caso deseje receber os resultados da pesquisa. Obrigada!

Este conteúdo não é criado nem endossado pela Microsoft. Os dados que você enviar serão enviados ao proprietário do formulário.